AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

© 2020 American Psychological Association
ISSN: 0096-3445

2021, Vol. 150, No. 5, 972–982
http://dx.doi.org/10.1037/xge0000987

She Told Me About a Singing Cactus: Counterintuitive Concepts Are More Accurately Attributed to Their Speakers Than Ordinary Concepts

Spencer Mermelstein
University of California, Santa Barbara

Michael Barlev
Arizona State University

Tamsin C. German
University of California, Santa Barbara

Communication is central to human life, yet it leaves humans vulnerable to misinformation and manipulation. Humans have therefore evolved a suite of psychological mechanisms for the evaluation of speakers and their messages. Here, we test a key hypothesized function of these "epistemic vigilance" mechanisms: the selective remembering of links between speakers and messages that are inconsistent with preexisting beliefs. Across four experiments, participants (N = 707) read stories associated with different contexts, with each story containing concepts that violate core knowledge intuitions ("counterintuitive concepts") and ordinary concepts. Experiment 1 revealed that after a brief distractor (2 min) participants more accurately attributed counterintuitive concepts to their speakers than ordinary concepts. Experiments 2a and 2b replicated this finding and found that this attribution accuracy advantage also extended to counterintuitive versus ordinary concepts associated with other contextual details—places and dates. Experiment 3 then tested whether this attribution accuracy advantage was more stable over time for speakers than for places. After a short distractor (20 min), there was a counterintuitive versus ordinary concept attribution accuracy advantage for both speakers and places. However, when participants were tested again after a long delay (48 hr), this attribution accuracy advantage more than doubled for speakers but disappeared entirely for places. We discuss the implications of these findings to the set of psychological mechanisms theorized to monitor and evaluate communication to guard our database of beliefs about the world.

Keywords: epistemic vigilance, communication, counterintuitive concepts, misinformation, source memory

Supplemental materials: http://dx.doi.org/10.1037/xge0000987.supp

Communication is central to human life. Human social living, from coordinating collective action to negotiating reciprocal exchange to social learning, is made possible by communication. Indeed, by allowing access to information stored in the minds of others, communication is an engine behind the evolution of cumulative cultural adaptations that cannot be discovered individually, from the processing of toxic plants for safe consumption or the Inuit's cold weather clothing (Henrich & McElreath, 2003), to the printing press, electricity, or semiconductors. Moreover, in human evolutionary history, communication obviated the need for individual trial-and-error learning in domains where errors can be catastrophic, such as learning which animals are dangerous (Barrett & Broesch, 2012) or which plants are edible (Wertz & Wynn, 2014). In short, communication is a key human adaptation that underlies our species' success across diverse ecologies.

However, because people vary in both their competence and trustworthiness, relying on communication opens listeners up to being misinformed or deceived (Sperber et al., 2010). Humans have therefore evolved a suite of psychological adaptations for epistemic vigilance or the evaluation of speakers and their messages (Sperber et al., 2010). Collectively, epistemic vigilance mechanisms guard our database of beliefs about the world.

The origins of such psychological adaptations have been observed in young children. As early as 2 years of age, toddlers not only update their beliefs in light of testimony from adults (Harris & Lane, 2014), but they also show sensitivity to the logical structure of arguments. For instance, toddlers are less likely to accept an argument backed by circular logic (e.g., "It's a fish,

This article was published Online First October 15, 2020.

Spencer Mermelstein, Department of Psychological and Brain Sciences, University of California, Santa Barbara; Michael Barlev, Department of Psychology, Arizona State University; Tamsin C. German, Department of Psychological and Brain Sciences, University of California, Santa Barbara.

Previous drafts of this article are available online (https://psyarxiv.com/6cp8e). The data reported here were presented in a talk at the May, 2019 Meeting of the Human Behavior and Evolution Society in Boston, Massachusetts. All data and materials are publicly available on the Open Science Framework (https://osf.io/x5k2u/).

Correspondence concerning this article should be addressed to Spencer Mermelstein, Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA 93106-9660. E-mail: mermelstein@psych.ucsb.edu

because I saw that it is a fish") compared with one based on evidence (e.g., "It's a fish, because I saw it swimming in the water"; Castelain, Bernard, & Mercier, 2018). By 5 years of age, children adjust their acceptance of messages on the basis of characteristics of speakers such as whether their previous testimony turned out to be true or false (e.g., Jaswal & Neely, 2006; Koenig & Harris, 2005), whether they were nice or mean to others in the past (e.g., Landrum, Mills, & Johnston, 2013; Mascaro & Sperber, 2009), and whether their previous testimony conformed with or dissented from a group consensus (e.g., Corriveau, Fusaro, & Harris, 2009; for a review see Harris, Koenig, Corriveau, & Jaswal, 2018).

The study of epistemic vigilance mechanisms takes on particular urgency in the present day as "fake news," political disinformation, and conspiracy theories proliferate on online platforms and elsewhere in a heretofore unprecedented scale (Lazer et al., 2018). For instance, long-impugned claims about a link between vaccinations and autism spectrum disorders fuel an "Anti-Vax" movement responsible for a worldwide reemergence of life-threatening infectious diseases (e.g., Larson, Cooper, Eskola, Katz, & Ratzan, 2011; Poland & Spier, 2010), and misinformation about anthropocentric global climate change reduces public support for mitigation efforts (e.g., Cook, Ellerton, & Kinkead, 2018; van der Linden, Leiserowitz, Rosenthal, & Maibach, 2017). A more thorough understanding of how epistemic vigilance mechanisms function could inform efforts to combat the proliferation and impact of such messages.

Here, we test experimentally a key hypothesized function of epistemic vigilance mechanisms: the selective remembering of links between speakers and messages that are inconsistent with preexisting beliefs (Sperber, 1997; Sperber et al., 2010; see also Mercier, 2017). By linking messages that violate preexisting beliefs to their speakers, listeners may continue evaluating these messages in light of new information about their speakers, as well as update their judgments about the speakers given new information about their messages. As an example, consider a scientist who makes a surprising claim about the dangers of a new vaccine that you believed to be safe. You view this scientist as trustworthy and competent, so you tentatively accept his claim. But your epistemic vigilance mechanisms might link the claim ("this new vaccine is dangerous") with its speaker (this scientist) for further evaluation of both the claim and the speaker. Should, in the future, factual errors with this claim be found, you may reevaluate its truth value (you might now reject the claim that this new vaccine is dangerous) as well as update your judgment about its speaker (you might now view the scientist as less competent and/or trustworthy). On this account, we expect a particularly robust link between speakers and, not everything they say, but specifically their messages that are inconsistent with preexisting beliefs.

As a secondary prediction, we test whether epistemic vigilance mechanisms monitor additional contextual details, or "meta-data", surrounding the acquisition of messages that violate prior beliefs (Cosmides & Tooby, 2000; Johnson, Hashtroudi, & Lindsay, 1993; Mahr & Csibra, 2017). For instance, Mahr and Csibra (2017) recently articulated a functional view of episodic memory wherein social interactions, in particular communicative exchanges, are remembered along with a set of contextual details such as the social background of the interaction (e.g., whether it happened in front of a group), when it happened, including relative to other events, and where it happened. Memory of such contextual details may further facilitate the ongoing evaluation of messages that violate preexisting beliefs. Returning to the above example, should we learn that the scientist who made this claim was at the time on the payroll of a rival vaccine company, we might doubt the accuracy of his claim more so than if the scientist had started working for this rival company a while after he made this claim.

However, two considerations suggest that links between messages that violate prior beliefs and their speakers, more so than links with other contextual details, should be of particular relevance to epistemic vigilance mechanisms. First, the truth value assigned to a message greatly depends on information about its speaker, generally more so than on other contextual details like the place or time of communication. For example, whether a message is accepted or rejected can entirely depend on whether its speaker is trustworthy or not. Second, messages reveal important information about their speakers such that linking messages to their speakers also allows listeners to update their judgment of these speakers should new information about their messages come to light. Thus, links between messages that violate preexisting beliefs and their speakers may be especially memorable as compared with such links with other contextual details.

We chose counterintuitive concepts as our test case of messages that violate preexisting beliefs. Counterintuitive concepts violate intuitions such as about folk physics, biology, and psychology (so-called core knowledge intuitions). For example, beliefs about people that can walk through walls violate intuitions about the solidity and spatiotemporal continuity of bodies (Boyer, 2001). As core knowledge intuitions reliably develop and are universally held (e.g., Carey, 2009), counterintuitive concepts are one class of communicated information that should be flagged by the epistemic vigilance mechanisms of listeners broadly as requiring further monitoring and evaluation.

In summary, we predicted better memory for the links between speakers, and potentially also other associated contextual details, and counterintuitive concepts as compared with ordinary concepts (those concepts consistent with prior beliefs).

Previous Research

Two literatures inform the present investigation. First, an extensive literature finds memory advantages for information that is inconsistent with preexisting beliefs (Hunt, 1995; von Restorff, 1933), such as information that violates stereotypes about social groups (e.g., Stangor & McMillan, 1992), schematic expectations (e.g., Hirshman, Whelley, & Palij, 1989; McDaniel & Einstein, 1986), and core knowledge intuitions (e.g., Banerjee, Haque, & Spelke, 2013; Barrett & Nyhof, 2001; Boyer & Ramble, 2001; Norenzayan, Atran, Faulkner, & Schaller, 2006). Erdfelder and Bredenkamp (1998) suggest that belief- or expectation-violating information differentially recruits attention, and as such undergoes more elaborate encoding that facilitates its later retrieval.

Second, studies from the literature on source memory find that schematic expectations or stereotypes bias memory judgments about the speakers or other contextual details associated with particular messages (Bayen, Nakamura, Dupuis, & Yang, 2000; Kuhlmann, Vaterrodt, & Bayen, 2012; Marsh, Cook, & Hicks, 2006; Mather, Johnson, & De Leonardis, 1999). For example, Bayen et al. (2000) found that utterances characteristic of medical doctors (e.g., "We are ready to run some tests"), yet spoken by a lawyer, were later misattributed to a doctor; Mather et al. (1999) found that utterances characteristic of Democrats (e.g., "I am pro-choice"), yet spoken by a Republican, were later misidentified as having been spoken by a Democrat. It has been suggested that such misattributions are a result of schema-based guessing biases: When participants cannot remember the speaker or other contextual details of a particular message, they select those that are schematically most likely to have been associated with it (e.g., Kuhlmann et al., 2012).

Nonetheless, other studies find that stimuli and the contextual details associated with them are better remembered when the stimuli are paired with an unexpected versus an expected context (Bell, Buchner, Kroneisen, & Giang, 2012; Ehrenberg & Klauer, 2005; Küppers & Bayen, 2014). For example, Küppers and Bayen (2014) presented participants with a word describing a particular location (e.g., "kitchen" or "bathroom") followed by items that were either schematically expected or unexpected of that location (e.g., "oven" or "toothbrush"). During a later memory task, participants were presented with the previously shown items and were asked to identify the location each item was paired with. Participants in this study were better at recalling locations that were unexpected for the items (e.g., "toothbrush" paired with "kitchen") compared with those that were expected for the items (e.g., "oven" paired with "kitchen"), which suggests that a violation of an expectation about the context with which an item is typically associated may enhance memory for that item-context pair.

Although previous studies have investigated memory for stimuli that violate prior beliefs (e.g., Boyer & Ramble, 2001) and memory for stimuli and their associated contexts when the pairing violates expectations (e.g., a toothbrush paired with a kitchen context; Küppers & Bayen, 2014), the present study is the first to explore memory for links between stimuli that by themselves violate preexisting beliefs and their associated contexts. With such a design we test a key hypothesized function of epistemic vigilance mechanisms concerning the "meta-data" (such as the associated speakers, places, and times) stored along with messages that violate preexisting beliefs, independent of any expectations about links between such messages and their speakers or when or where the information was communicated.

Counterintuitive Concepts

The human conceptual repertoire is founded in part on species-typical, reliably developing core knowledge mechanisms that are specialized for representing concepts from domains such as physical objects and their spatiotemporal properties and mechanics ("folk physics"), human-made artifacts including tools, animals and their biology ("folk biology"), plants, and persons and their mental states ("folk psychology"; e.g., Baillargeon, Scott, & Bian, 2016; Barrett et al., 2013; Carey, 2009; German & Barrett, 2005; Inagaki & Hatano, 2002; Spelke, 1990; Spelke & Kinzler, 2007; Wertz, 2019). For example, infants understand that objects are cohesive and bounded wholes that neither separate nor coalesce, and that objects only move on contact (Baillargeon, 2004; Spelke, Breinlinger, Macomber, & Jacobson, 1992). Infants also interpret and predict the behavior of persons in terms of internal mental states, understand that beliefs are linked to perceptions, and that people can have beliefs that are false (Baillargeon et al., 2016; Onishi & Baillargeon, 2005).

The human mind is also capable of representing concepts that violate core knowledge intuitions—indeed, such counterintuitive concepts are widespread in science (Shtulman, 2017) and religion (Boyer, 1994, 2001, 2003). For example, the concept of heritable genetic mutations, a fundamental principle of the theory of evolution by natural selection, violates folk biological intuitions about the immutability of animal "essences"; the concept of a statue that can hear prayers violates folk physical intuitions by transferring a psychological property to a human-made artifact. Although counterintuitive concepts violate universal intuitions, people nonetheless come to believe in many such concepts and may even hold in high esteem those with expertise about these concepts (e.g., scientists and religious specialists).

Barlev and colleagues (Barlev, Mermelstein, Cohen, & German, 2019; Barlev, Mermelstein, & German, 2017, 2018) recently presented empirical evidence that even though counterintuitive concepts are widely believed in, they cannot be fully reconciled with the core knowledge intuitions with which they conflict (also see Barrett, 1998; Barrett & Keil, 1996). The God concept in Christianity, for example, is initially built by coopting the person "template," a set of core intuitions about the physical, biological, and psychological properties of people. Accordingly, young children reason that God is capable of having beliefs that are false just like persons, and it is only later that children come to view God (but not ordinary people) as infallible (Lane, Wellman, & Evans, 2010). Barlev and colleagues used a statement verification task where adult Christian religious adherents evaluated as "true" or "false" statements that were consistent or inconsistent between core intuitions about persons and acquired theology about God. As predicted, participants were slower and less accurate at verifying inconsistent statements as compared with consistent statements, suggesting that core knowledge intuitions about the psychology (Barlev et al., 2017, 2018) and physicality (Barlev et al., 2019) of persons coexist and interfere with acquired beliefs about God (e.g., infallibility).

Thus, counterintuitive concepts are an ideal case for testing predictions about the functioning of epistemic vigilance mechanisms: because counterintuitive concepts violate, and cannot be reconciled with, universally held core knowledge intuitions, they should be flagged by the epistemic vigilance mechanisms of listeners broadly as warranting further monitoring and evaluation.

The Current Study

Across four experiments, participants read a series of short stories, with each story containing counterintuitive and ordinary concepts, and each story transmitted by different persons (Experiments 1 through 3) or at different places (Experiments 2a and 3) or on different dates (Experiment 2b). Then, following a delay, participants were asked to attribute each concept to its associated context. Given our goal of investigating memory for links between messages that violate prior beliefs and the context of their acquisition, it was critical that we presented each speaker, place, or date with an equal number of counterintuitive and ordinary concepts. Without this design feature, attributions made during the task might be regulated not by remembered links between specific concepts and their speakers, but by general associations formed between some speakers with counterintuitive concepts and other speakers with ordinary concepts.

Experiment 1 tested the prediction that counterintuitive concepts are more accurately attributed to their speakers than ordinary concepts. Experiments 2a and 2b replicated this and tested whether counterintuitive concepts associated with different contextual details, places (Experiment 2a), and times (Experiment 2b), also exhibit a counterintuitive versus ordinary concepts attribution accuracy advantage. Last, Experiment 3 used longer periods of delay, with a first attribution phase after a 20-min delay and a second attribution phase after a 48-hr delay, to examine the relative stability of the links between concepts and their associated contextual details. We predicted that counterintuitive versus ordinary concepts would exhibit an attribution accuracy advantage, and that this effect would be more stable over time for speakers than for other contextual details.

Experiment 1

The goal of Experiment 1 was to test whether counterintuitive concepts are more accurately attributed to their speakers than ordinary concepts.

Method

Participants. A priori power analyses were computed for all experiments reported here (see the online supplemental material). Participants (N = 107; 66% female) were undergraduates at the University of California, Santa Barbara (UCSB; M_{age} = 19.4, SD = 2.27), who in this and all other experiments reported here received course credit for their participation. Participants identified as East, South, or Southeast Asian (35%); White (32%); Hispanic or Latino (22%); or as another ethnic/racial background (11%). All experiments in this article were approved by UCSB's Institutional Review Board (Protocol 23–18-0027), and informed consent was obtained from all participants.

Design. The independent variable was concept (counterintuitive vs. ordinary), presented within-subjects. The dependent variables were the proportion of counterintuitive and ordinary concepts correctly attributed to their speaker.

Materials and procedure. Materials were adapted from Banerjee et al. (2013) and consist of four 340-word stories, each associated with a different speaker, and each containing three counterintuitive and three ordinary concepts (for a total of 24 concepts across the four stories). The concepts were created as follows. Three pairs of nouns (e.g., *cat–dog*) were generated in each of the following domains: animals, plants, nonliving natural objects, and human-made artifacts. Each noun was embedded in a descriptor composed of two adjectival clauses: a first clause that is consistent with the domain and a second clause that is either also consistent (forming an ordinary concept) or contains a violation of a physical, biological, or psychological core knowledge intuition held about the domain (forming a counterintuitive concept). For example, the noun *cat* was paired with either the ordinary descriptor "has soft fur and likes to play with toys" or the counterintuitive descriptor "has brown spots and can walk through solid walls" (a violation of intuitive physics). The two variants of each concept (Cat–Dog + Ordinary Descriptor and Cat–Dog + Counterintuitive Descriptor) were controlled for number of words per sentence and were balanced in terms of overall sentence structure and complexity. See Table 1 for sample concepts. See the online supplemental material for all concepts.

Two lists of concept stimuli were created by varying which descriptor (counterintuitive or ordinary) was linked with which noun in a pair. For example, in List 1, *cat* was paired with the counterintuitive descriptor (and *dog* was paired with the ordinary descriptor), whereas in List 2, *cat* was paired with the ordinary descriptor (and *dog* was paired with the counterintuitive descriptor). Participants were randomly assigned one of the two concept stimuli lists such that, between lists, the descriptors remained fixed but the noun that they were paired with was varied. In this way, we could verify that attribution accuracy was a function of the type of descriptor (counterintuitive or ordinary), rather than a property of specific noun–descriptor pairings.

Participants were asked to imagine that they frequently go camping with four close friends named Miguel, Joanna, Sam, and Ariel, and that during one of these trips, each friend took a turn telling the participant one of the four short stories. Critically, to prevent participants from broadly associating certain types of concepts to certain speakers, three ordinary and three counterintuitive concepts were randomly distributed throughout the middle of each story, such that each friend was associated with an equal number of both types of concepts.

Finally, participants were randomly assigned to receive one of four different versions of the task, created by varying which person was associated with which story, such that each person was associated with each story across the different versions. Task Versions 1 and 2 used Stimuli List 1 and Task Versions 3 and 4 used Stimuli List 2. See the following text for an example

Table 1
Example Counterintuitive and Ordinary Concepts

		Descriptor	
Noun pair	Domain	Counterintuitive	Ordinary
Cat–dog	Animal	Has brown spots and can walk through solid walls	Has soft fur and likes to play with toys
Shrub–cactus	Plant	Is small and likes to sing loudly	Is dark green and grows next to streams
Branch–rock	Object	Is cold to the touch and can speak in French	Is thick and hard and looks shiny in the sunlight
Table–chair	Artifact	Is big and often floats in midair	Is firm to the touch and can hold lots of weight

Note. Counterintuitive descriptors contain violations of core knowledge intuitions. Concepts are modified from those in Banerjee, Haque, and Spelke (2013).

of one of the short stories and the online supplemental material for all stories.

> [Miguel/Joanna/Sam/Ariel] tells you the following story:
>
> A brother and a sister moved with their parents to a new house on a new street that they had never seen before. The new house was in a neighborhood several miles away from where they used to live. The brother and sister were excited to explore their new home and to learn more about the neighborhood. As soon as their boxes were unpacked, the brother and sister decided to go see what they could find in and around their new home.
>
> First, they climbed up a staircase and went into the attic, where they saw a lizard on the floor. This was a lizard that had a long, thin tail and could never die no matter what happened to it. The kids left the attic and wandered to their parent's bedroom. In the bedroom, they saw a hammer lying on the carpet. The hammer had a wooden handle and needed food every day to stay strong. After leaving the bedroom, the kids continued into the basement, where they noticed a shovel on top of a table. The shovel felt heavy to hold and was a light brown in color.
>
> Growing bored of the house, the kids went outdoors into their new backyard. They looked up and saw a rainbow. This rainbow was high in the sky and could be seen from the ground. The kids skipped down the street and came across a garden that had a single rose in it. The rose swayed in the wind and could be in two different parts of the world at the exact same time. The kids finally reached the front yard of their closest neighbor's house. On the lawn, the kids spotted a rat. The rat ate insects off the ground and moved around quickly on all four of its feet.
>
> Satisfied with what they had seen, the kids went back inside thinking that their new home was going to be a very interesting place to live.

Participants were tested in groups of up to eight in semiprivate computer workstations. Qualtrics software was used to administer all experiments. Data were analyzed using RStudio 3.5.1 (RStudio Team, 2020) and JASP 0.9 (JASP Team, 2017). (Qualtrics scripts, data, and R code are available on the Open Science Framework at the link provided in the author note.) Participants were instructed to "pay particularly careful attention to the person who is telling you the story and what happens in the story" and that they would "need to remember this information for a memory test that will occur later in the study." During the encoding phase, the stories were presented one at a time and in a random order. Each story was "locked" on the screen for 90 s (estimated as the average reading time across the four stories), after which participants were allowed to continue whenever they were ready; this was done to ensure that participants did not speed through the stories. After reading each story, as a check that they have read that story and to verify that they encoded the person associated with the story, participants were asked to identify the friend who told them that story in a forced choice question. During the distractor phase—lasting 2 min in this experiment—participants were shown a blank map of the United States and were asked to type the names of as many states as they could. Last, during the attribution phase, participants were presented with the 24 concepts they read during the encoding phase, one at a time and in randomized order, along with the names of the four friends with whom the concepts were associated. Participants were instructed to "identify, as accurately as possible, which of your friends was the one who told you each statement." The entire study took approximately 20 min. Figure 1 summarizes this procedure.

Results

In this and all other experiments reported in this article there were no statistically significant differences between stimuli lists or task versions. A paired-samples *t* test revealed, as predicted, that counterintuitive concepts were more accurately attributed to their speakers than ordinary concepts, $t(106) = 5.05$, $p < .001$, $d = 0.49$, 95% CI [0.29, 0.69]. See Figure 2 for a pirate plot.

Experiment 2a

As predicted, Experiment 1 found that after a brief delay counterintuitive concepts were more accurately attributed to their speakers than ordinary concepts. The goal of Experiment 2a-b was to investigate whether other contextual details also show an attribution accuracy advantage for counterintuitive versus ordinary

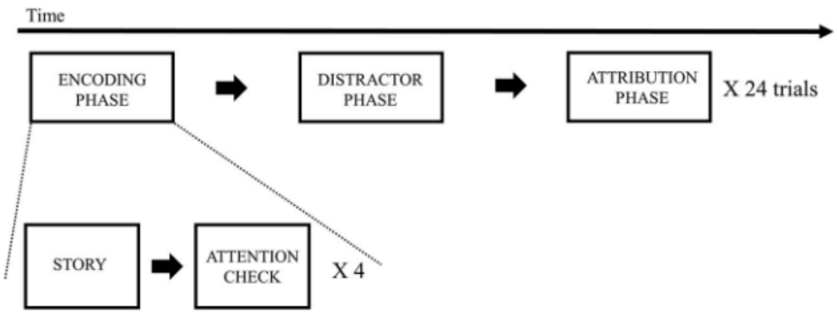

Figure 1. Summary of the procedure used in Experiments 1 through 3. Participants read four 340-word stories, each containing three counterintuitive and three ordinary concepts, and each associated with a different speaker or other contextual information (places or dates). After reading each story, participants completed an attention check to verify they read and remembered the speaker or other contextual information. In Experiment 1, Experiment 2a, and Experiment 2b, there was a distractor phase lasting 2 min before the attribution phase, where participants were asked to attribute each concept to the speaker or context with which it was associated. In Experiment 3, there were two attribution phases, one after a distractor phase lasting 20 min and another after a 48-hr delay.

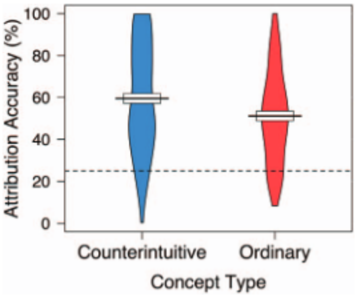

Figure 2. Pirate plot of mean attribution accuracy (%) for counterintuitive and ordinary concepts in Experiment 1. Inference bands correspond to 95% within-subjects confidence intervals. The dotted line at 25% indicates chance performance. See the online article for the color version of this figure.

concepts. In doing so, we tested between two alternative possibilities of what "meta-data" is linked to messages that violate preexisting beliefs. As we argued above, links between messages that violate preexisting beliefs and their speakers are plausibly more relevant to epistemic vigilance mechanisms than links between such messages and other contextual details. Thus, one possibility is that the attribution accuracy advantage for contextual details like places and dates would be smaller as compared with persons. Alternatively, it is nonetheless possible that a broad variety of metadata remains linked to messages that violate preexisting beliefs. On this account, after a brief delay, speakers and contextual details such as where or when a message was acquired will show a similar counterintuitive versus ordinary concepts attribution accuracy advantage. Experiments 2a and 2b tested between these two accounts by comparing the attribution accuracy of counterintuitive versus ordinary concepts linked with speakers versus places (Experiment 2a) and speakers versus dates (Experiment 2b).

Method

Participants. Participants were 200 (64% female) UCSB undergraduates ($M_{\text{age}} = 18.9$, $SD = 1.23$). Participants identified as White (40%); East, South, or Southeast Asian (30%); Hispanic or Latino (20%); or as another ethnic/racial background (10%).

Design. This study used a 2 (concept: counterintuitive vs. ordinary) × 2 (condition: person vs. place) design with repeated measures on the first factor. The dependent variables were the proportions of counterintuitive and ordinary concepts correctly attributed to their associated person or place.

Materials and procedure. Participants were randomly assigned to the person or place condition. The person condition was identical to Experiment 1. In the place condition, instead of information about a speaker, each story began with information about a national park where the story was told ("While you are camping in [Mammoth/Big Sur/Joshua Tree/Sequoia], you hear the following story"). The rest of the procedure was the same as in Experiment 1, except participants in the place condition were asked to attribute each concept to the place where they were told about it.

Results

Attribution accuracy means were entered into a 2 (concept: counterintuitive vs. ordinary) × 2 (condition: person vs. place) mixed analysis of variance (ANOVA) with repeated measures on the first factor. Results revealed a main effect of concept, $F(1, 198) = 29.36$, $p < .001$, $\eta_p^2 = .13$, no main effect of condition, $F(1, 198) < 1.0$, $p > .250$, and no interaction between the two, $F(1, 198) = 1.25$, $p > .250$. After a brief delay, counterintuitive concepts were more accurately attributed to their associated persons or places than ordinary concepts, and this effect was not statistically different for persons as compared with places. See Figure 3 for pirate plots.

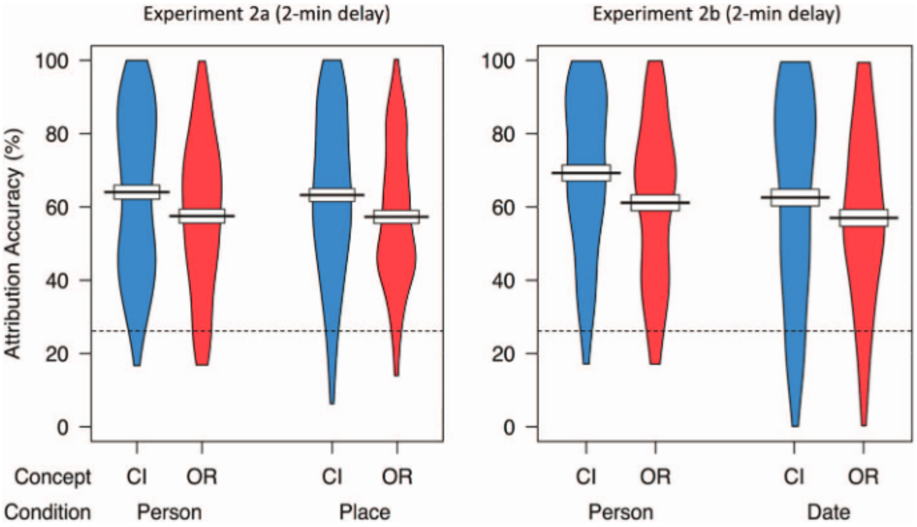

Figure 3. Pirate plots of mean attribution accuracy (%) for counterintuitive (CI) and ordinary (OR) concepts in Experiments 2a and 2b. Inference bands correspond to 95% within-subjects confidence intervals. The dotted line at 25% indicates chance performance. See the online article for the color version of this figure.

Experiment 2b

Method

Participants. Participants were 188 (78% female) UCSB undergraduates (M_{age} = 18.9, *SD* = 1.13). Participants identified as East, South, or Southeast Asian (36%); White (29%); Hispanic or Latino (25%); or as another ethnic/racial background (10%).

Design. This study used a 2 (concept: counterintuitive vs. ordinary) × 2 (condition: person vs. date) design with repeated measures on the first factor. The dependent variables were the proportions of counterintuitive and ordinary concepts correctly attributed to their associated persons or dates.

Materials and procedure. Participants were randomly assigned to the person or date condition. The person condition was identical to Experiment 1. In the date condition, instead of information about a speaker, each story began with information about a date on which the story was told ("On [April 7/April 12/April 19/April 26] a friend tells you the following story"). The rest of the procedure was the same as in Experiment 1, except participants in the date condition were asked to attribute each concept to the date on which they were told about it.

Results

Attribution accuracy means were entered into a 2 (concept: counterintuitive vs. ordinary) × 2 (condition: person vs. date) mixed ANOVA with repeated measures on the first factor. Results revealed a main effect of concept, $F(1, 186) = 24.59$, $p < .001$, $\eta_p^2 = .12$, no main effect of condition, $F(1, 186) = 2.44$, $p = .120$, and no interaction between the two, $F(1, 186) < 1.0$, $p > .250$. After a brief delay, counterintuitive concepts were more accurately attributed to their associated speakers or dates than ordinary concepts, and this effect was not statistically different for persons as compared with dates. See Figure 3 for pirate plots.

Experiment 3

Experiments 2a and 2b replicated and extended Experiment 1 by demonstrating that after a brief delay counterintuitive versus ordinary concepts were more accurately attributed not only to their speakers, but also to other contextual details: their places and times of acquisition. The attribution accuracy advantage for counterintuitive versus ordinary concepts in these experiments was not statistically different for speakers as compared places or dates, suggesting that epistemic vigilance mechanisms may initially flag a variety of contextual details surrounding the acquisition of messages that violate preexisting beliefs. We next explored the stability over time of links between such messages and their associated contextual details.

In Experiment 3, participants completed the attribution task twice, once after a short distractor phase (20 min) and again after a 48-hr delay. We predicted that counterintuitive concepts would be more accurately attributed to the contexts of their acquisition than ordinary concepts, and that this advantage would be more stable over time for speakers as compared with other contextual details.

Method

Participants. Participants were 212 (73% female) UCSB undergraduates (M_{age} = 18.7, *SD* = 1.09). Participants identified as White (40%); East, South, or Southeast Asian (28%); Hispanic or Latino (24%); or as another ethnic/racial background (8%). Of these, 194 (92%) returned for the second session. We report results from participants who completed both sessions only. The pattern of results for the first session remains the same if we analyze data from the full sample.

Design. This study used a (concept: counterintuitive vs. ordinary) × 2 (delay: 20 min vs. 48 hr) × 2 (condition: person vs. place) design with repeated measures on the first two factors. The dependent variables were the proportions of counterintuitive and ordinary concepts correctly attributed to their associated persons or places.

Materials and procedure. The procedure was identical to that in Experiment 2a, except that after the encoding task, participants completed a 20-min (rather than a 2-min) battery of distractor tasks before the first attribution task. After 48 hr, participants then returned for a second testing session to complete the attribution task again. Although participants knew there would be a second session, they were not told they would be tested for their memory of the first session stimuli again.

Results

Attribution accuracy means were entered into a 2 (concept: counterintuitive vs. ordinary) × 2 (delay: 20 min vs. 48 hr) × 2 (condition: person vs. place) mixed ANOVA with repeated measures on the first two factors. Results revealed a main effect of concept, $F(1, 192) = 24.69$, $p < .001$, $\eta_p^2 = 0.11$, a main effect of delay, $F(1, 192) = 69.39$, $p < .001$, $\eta_p^2 = 0.27$, and no main effect of condition, $F(1, 192) < 1.0$, $p > .250$. There were no two-way interactions: Concept × Delay, $F(1, 192) < 1.0$, $p > .250$, Condition × Delay, $F(1, 192) < 1.0$, $p > .250$, and Condition × Concept, $F(1, 192) = 3.67$, $p = .057$. Critically, there was a three-way Concept × Delay × Condition interaction, $F(1, 192) = 10.36$, $p = .002$, $\eta_p^2 = 0.05$. We unpack this three-way interaction below. See Figure 4 for pirate plots and online supplemental material for an alternative analytic approach using difference scores. Both approaches yielded the same conclusions.

Attribution accuracy advantage for persons versus places after 20 min and 48 hr. After a 20-min delay, the counterintuitive versus ordinary concepts attribution accuracy advantage did not statistically differ between persons and places, $t(192) < 1.0$, $p > .250$, thereby replicating the findings of Experiment 2a. However, after a 48-hr delay, this attribution accuracy advantage was significantly greater for persons as compared with places, $t(192) = 3.46$, $p < .001$, $d = .50$, 95% CI [0.21, 0.78].

Change in attribution accuracy over time for persons and places. Simple main effect analyses evaluated attribution accuracy for counterintuitive versus ordinary concepts over time, separately in the person and place conditions. In the person condition, the attribution accuracy advantage for counterintuitive versus ordinary concepts more than doubled with time: after 20 min, $t(99) = 2.54$, $p = .012$, $d = 0.26$, 95% CI [0.05, 0.45]; after 48 hr, $t(99) = 5.68$, $p < .001$, $d = 0.57$, 95% CI [0.36, 0.78]. In the place condition, the attribution accuracy advantage for counterintuitive versus ordinary concepts disappeared entirely with time: after 20

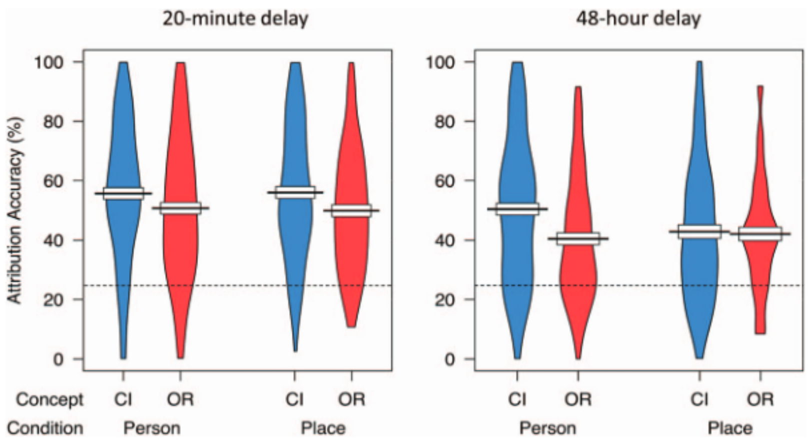

Figure 4. Pirate plots of mean attribution accuracy (%) for counterintuitive (CI) and ordinary (OR) concepts after 20 min and 48 hr. Inference bands correspond to 95% within-subjects confidence intervals. The dotted line at 25% indicates chance performance. See the online article for the color version of this figure.

min, $t(93) = 2.92$, $p = .004$, $d = 0.30$, 95% CI [0.09, 0.51]; after 48 hr, $t(93) < 1.0$, $p > .250$.

Comparing rates of decline in attribution accuracy over time. Attribution accuracy for person-counterintuitive concepts (CI) pairs started higher than that for person-ordinary concepts (OR) pairs (M_{CI} = 55.7% vs. M_{OR} = 50.4%) and was more stable over time ($M_{\text{difference}} = -5.2\%$, $SE_{\text{difference}} = 1.9\%$ vs. $M_{\text{difference}} = -10.7\%$, $SE_{\text{difference}} = 1.6\%$, respectively; $t[99] = 2.55$, $p = .012$, $d = 0.26$, 95% CI [0.06, 0.45]). Attribution accuracy for person-CI pairs started about the same as for place-CI pairs (M_{CI} = 54.1%) but was more stable than it over time ($M_{\text{difference}} = -5.2\%$, $SE_{\text{difference}} = 1.9\%$ vs. $M_{\text{difference}} = -11.3\%$, $SE_{\text{difference}} = 1.9\%$, respectively; $t[192] = 2.27$, $p = .025$, $d = 0.33$, 95% CI [0.04, 0.61]). On the other hand, attribution accuracy for place-CI pairs started higher than for place-OR pairs (M_{CI} = 54.1% vs. M_{OR} = 42.7%) but was less stable over time ($M_{\text{difference}} = -11.3\%$, $SE_{\text{difference}} = 1.9\%$ vs. $M_{\text{difference}} = -6.5\%$, $SE_{\text{difference}} = 2.0\%$, respectively; $t[93] = 2.03$, $p = .045$, $d = 0.21$, 95% CI [0.004, 0.41]). There was no significant difference in attribution accuracy over time for person-OR versus place-OR pairs, $t(192) = -1.70$, $p = .090$, $d = -0.25$, 95% CI [−0.53, 0.04].

In sum, after a 20-min delay, there was an attribution accuracy advantage for counterintuitive versus ordinary concepts associated with persons or with places, and the two did not statistically differ. However, after 48-hr, this attribution accuracy advantage more than doubled in size for persons; this was due to the relative stability of attribution accuracy for person-CI links over time as compared with person-OR links. Conversely, the counterintuitive versus ordinary concepts attribution accuracy advantage for places disappeared entirely after 48-hr; this was due to a relatively rapid decline of attribution accuracy over time for place-CI links as compared with place-OR links.

General Discussion

Communication is central to human life. Yet communication leaves listeners vulnerable to misinformation and manipulation. Consequently, it has been proposed that humans evolved a suite of adaptations—collectively termed *epistemic vigilance mechanisms*—to mitigate such threats by monitoring and evaluating communication (Sperber et al., 2010). Here, we tested the hypothesis that epistemic vigilance mechanisms selectively remember the links between speakers and messages that are inconsistent with preexisting beliefs (Sperber, 1997; Sperber et al., 2010; see also Cosmides & Tooby, 2000; Johnson et al., 1993). We tested this hypothesis using the case study of concepts that violate core knowledge intuitions about folk physics, biology, and psychology (counterintuitive concepts; e.g., Boyer, 2001). Across four experiments, participants read stories containing counterintuitive concepts (e.g., “a cat that has brown spots and can walk through solid walls”) and ordinary concepts (e.g., “a dog that has soft fur and likes to play with toys”) that were associated with persons or with other contextual details (places or times). After a delay, participants were asked to attribute these concepts to the context of their acquisition.

As predicted, Experiment 1 found that after a brief delay (2 min) participants were better at attributing counterintuitive than ordinary concepts to their speakers. Experiments 2a and 2b replicated these findings and further found that this attribution accuracy advantage for counterintuitive versus ordinary concepts extended to other contextual details: places (Experiment 2a) and dates (Experiment 2b). Thus, after a brief delay, a broad variety of contextual details are differentially linked in memory to messages that violate preexisting beliefs (e.g., Cosmides & Tooby, 2000; Johnson et al., 1993).

We hypothesized, however, that it may be especially relevant for epistemic vigilance mechanisms to remember who told you a message that is inconsistent with your preexisting beliefs, more so than where or when you heard this message. Given this, we explored the possibility that the links between messages that violate preexisting beliefs and their speakers are especially stable over time compared with links between such messages and other contextual details.

Experiment 3 tested this using repeated attribution tests. After a short distractor phase (20 min), participants were better at attributing counterintuitive than ordinary concepts to their associated contextual details, and this memory advantage did not differ for speakers versus places. After a 48-hr delay, however, participants no longer showed an attribution accuracy advantage for counterintuitive versus ordinary concepts and their associated places. In contrast, participants were not only still better at attributing counterintuitive versus ordinary concepts to their speakers, but this effect more than doubled.

The current study advances our understanding of how epistemic vigilance mechanisms monitor and evaluate communication. Epistemic vigilance mechanisms detect inconsistencies between acquired messages and preexisting beliefs, and selectively remember contextual details surrounding the acquisition of such messages, with memory for links between such messages and their speakers being especially stable over time. The linking of messages that violate preexisting beliefs with such meta-data is a key function of epistemic vigilance mechanisms, as they are then able to continue evaluating these messages should new information about the competence or trustworthiness of their speakers come to light, as well as continue evaluating speakers given new information about their messages.

Indeed, linking messages to metadata about their speakers is a plausible step toward developing profiles of our social partners as sources of information. Messages that are at odds with preexisting beliefs are particularly informative in this regard, as these could reveal that their speakers have information that we do not, or that they are incompetent or even deceptive. For instance, should one friend spread negative rumors that are at odds with your positive opinion of a mutual friend, your epistemic vigilance mechanisms might associate this claim with its speaker, and you might be motivated to search for additional information about the claim and/or its speaker as you attempt to reconcile the claim with your preexisting beliefs. Whether you subsequently accept or reject the claim, remembering the link between the claim and its speaker might still be advantageous, as it can influence your decisions on whether to believe future things that speaker says.

Moreover, our findings add to a growing literature (e.g., Mayo, 2019; Mercier, 2017, 2020) suggesting that, contrary to previous accounts, humans are not unduly gullible. Believing misinformation, such as "fake news," political propaganda, or conspiracies may instead mainly be a function of its fit (or lack thereof) with preexisting beliefs and motivations. Thus, as recommended by Lewandowsky, Ecker, Seifert, Schwarz, and Cook (2012), for example, targeting factors such as an audience's preexisting beliefs may be a productive starting point in combating the spread of misinformation.

The findings reported here are also relevant to the source memory literature. In contrast to past studies on source memory that leveraged violations of expectations about the pairing of stimuli and their associated contexts (e.g., a toothbrush paired with a kitchen setting vs. a bathroom setting; Küppers & Bayen, 2014), the studies reported here demonstrate that stimuli that violate preexisting beliefs by themselves are selectively linked to their contextual details.

Future research is needed to shed light on the exact mechanism by which links between messages that violate prior beliefs and their associated contexts are remembered. We consider it possible that epistemic vigilance mechanisms store such messages in a "metarepresentational" data structure that is specialized for linking messages to their metadata (Cosmides & Tooby, 2000). As suggested by Leslie (1987), metarepresentation constitutes the minimal cognitive architecture needed to decouple representations from one's existing database of beliefs, including representations of the mental states of others (mentalizing) and counterfactuals (e.g., pretend play). As metarepresentations, messages that violate preexisting beliefs are hypothesized to remain quarantined, along with metadata about the context of their acquisition, pending further evaluation (Mercier, 2017; Sperber, 1997; Sperber et al., 2010).

Alternatively, the mind might use other mechanisms to link messages that violate prior beliefs with their metadata. For instance, on recall people may reconstruct the links between speakers and their messages. In the experiments reported here, participants could have remembered who the speaker of, say, the first story presented was, and, independent of this, remembered the concepts that were in that first story, thereby allowing them to identify the speaker of the concepts in the first story. In other words, rather than a direct speaker-concept link, participants could have formed speaker-story and concept-story links that allowed them to reconstruct the speaker-concept link.[1] Regardless of the exact mechanism by which metadata about messages that violate prior beliefs is stored, the selective remembering of metadata surrounding the acquisition of such messages, as demonstrated in the experiments reported here, is a key predicted function of epistemic vigilance mechanisms, and facilitates their capacity for monitoring and evaluating communication.

Future research may also investigate the broader range of messages that trigger epistemic vigilance mechanisms. For example, messages that are improbable but not impossible, such as "there are alligators in the New York City sewers," are likely to be subjected to epistemic scrutiny by adults and also by children, who seem to have a weaker grasp of the improbable versus impossible distinction (Shtulman & Carey, 2007). Moreover, Sperber et al. (2010) suggest that epistemic vigilance mechanisms are sensitive to the personal relevance of a message. Thus, one might be more likely to scrutinize a claim about the existence of alligators in the New York City sewers if she lives in New York City as compared with California.

In conclusion, we demonstrated that people selectively remember the links between messages that violate preexisting beliefs and their contextual details, especially their speakers. Memory for the context in which messages that violate preexisting beliefs are shared may be crucial to the ongoing monitoring and evaluation of such messages, particularly the differentiation of beneficial from harmful messages, and to constructing profiles of our social partners. Human social living is made possible by communication, and in turn, it is psychological mechanisms like those studied here that safeguard us against misinformation and make communication advantageous.

Context of the Research

Communication is central to human social life, yet it exposes listeners to misinformation and manipulation. Here, we study the cognitive mechanisms that are theorized to have evolved to keep communication advantageous. We focus on one hypothesized function of these epistemic vigilance mechanisms: the representation of the contextual details—such as the speakers—of information that is inconsistent with preexisting beliefs. Our study was inspired by Sperber et al. (2010), who articulated the theoretical logic of these epistemic vigilance adaptations, and Sperber (1997), who suggested that messages that are inconsistent with preexisting beliefs are stored

[1] The authors thank Karen J. Mitchell for raising this possibility.

along with speaker tags. We were also inspired by Cosmides and Tooby (2000) who articulated a broader theoretical model of the meta-data stored along with messages. In our prior research (Barlev et al., 2017, 2018, 2019), we found that concepts that conflict with universally held core knowledge intuitions (counterintuitive concepts) do not revise those intuitions but coexist alongside them. We have therefore used counterintuitive concepts as our test case here. The present study is part of a broader research program into the functions of epistemic vigilance mechanisms and how they are used to critically evaluate communication and thereby guard our database of beliefs from misinformation.

References

Baillargeon, R. (2004). Infants' physical world. *Current Directions in Psychological Science, 13,* 89–94. http://dx.doi.org/10.1111/j.0963-7214.2004.00281.x

Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology, 67,* 159–186. http://dx.doi.org/10.1146/annurev-psych-010213-115033

Banerjee, K., Haque, O. S., & Spelke, E. S. (2013). Melting lizards and crying mailboxes: Children's preferential recall of minimally counter-intuitive concepts. *Cognitive Science, 37,* 1251–1289. http://dx.doi.org/10.1111/cogs.12037

Barlev, M., Mermelstein, S., Cohen, A. S., & German, T. C. (2019). The embodied God: Core intuitions about person physicality coexist and interfere with acquired Christian beliefs about God, the Holy Spirit, and Jesus. *Cognitive Science, 43,* e12784. http://dx.doi.org/10.1111/cogs.12784

Barlev, M., Mermelstein, S., & German, T. C. (2017). Core intuitions about persons coexist and interfere with acquired Christian beliefs about God. *Cognitive Science, 41*(Suppl. 3), 425–454. http://dx.doi.org/10.1111/cogs.12435

Barlev, M., Mermelstein, S., & German, T. C. (2018). Representational coexistence in the God concept: Core knowledge intuitions of God as a person are not revised by Christian theology despite lifelong experience. *Psychonomic Bulletin & Review, 25,* 2330–2338. http://dx.doi.org/10.3758/s13423-017-1421-6

Barrett, H. C., & Broesch, J. (2012). Prepared social learning about dangerous animals in children. *Evolution and Human Behavior, 33,* 499–508. http://dx.doi.org/10.1016/j.evolhumbehav.2012.01.003

Barrett, H. C., Broesch, T., Scott, R. M., He, Z., Baillargeon, R., Wu, D., . . . Laurence, S. (2013). Early false-belief understanding in traditional non-Western societies. *Proceedings of the Royal Society B: Biological Sciences, 280,* 20122654. http://dx.doi.org/10.1098/rspb.2012.2654

Barrett, J. L. (1998). Cognitive constraints on Hindu concepts of the divine. *Journal for the Scientific Study of Religion, 37,* 608–619. http://dx.doi.org/10.2307/1388144

Barrett, J. L., & Keil, F. C. (1996). Conceptualizing a nonnatural entity: Anthropomorphism in God concepts. *Cognitive Psychology, 31,* 219–247. http://dx.doi.org/10.1006/cogp.1996.0017

Barrett, J. L., & Nyhof, M. A. (2001). Spreading non-natural concepts: The role of intuitive conceptual structures in memory and transmission of cultural materials. *Journal of Cognition and Culture, 1,* 69–100. http://dx.doi.org/10.1163/156853701300063589

Bayen, U. J., Nakamura, G. V., Dupuis, S. E., & Yang, C.-L. (2000). The use of schematic knowledge about sources in source monitoring. *Memory & Cognition, 28,* 480–500. http://dx.doi.org/10.3758/BF03198562

Bell, R., Buchner, A., Kroneisen, M., & Giang, T. (2012). On the flexibility of social source memory: A test of the emotional incongruity hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1512–1529. http://dx.doi.org/10.1037/a0028219

Boyer, P. (1994). *The naturalness of religious ideas: A cognitive theory of religion*. Berkeley, CA: University of California Press. Retrieved from https://www.ucpress.edu/book/9780520075597/the-naturalness-of-religious-ideas

Boyer, P. (2001). *Religion explained: The evolutionary origins of religious thought*. New York, NY: Basic Books. Retrieved from https://www.basicbooks.com/titles/pascal-boyer/religion-explained/9780465004614/

Boyer, P. (2003). Religious thought and behaviour as by-products of brain function. *Trends in Cognitive Sciences, 7,* 119–124. http://dx.doi.org/10.1016/S1364-6613(03)00031-7

Boyer, P., & Ramble, C. (2001). Cognitive templates for religious concepts: Cross-cultural evidence for recall of counter-intuitive representations. *Cognitive Science, 25,* 535–564. http://dx.doi.org/10.1207/s15516709cog2504_2

Carey, S. (2009). *The origin of concepts*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195367638.001.0001

Castelain, T., Bernard, S., & Mercier, H. (2018). Evidence that two-year-old children are sensitive to information presented in arguments. *Infancy, 23,* 124–135. http://dx.doi.org/10.1111/infa.12202

Cook, J., Ellerton, P., & Kinkead, D. (2018). Deconstructing climate misinformation to identify reasoning errors. *Environmental Research Letters, 13,* 024018. http://dx.doi.org/10.1088/1748-9326/aaa49f

Corriveau, K. H., Fusaro, M., & Harris, P. L. (2009). Going with the flow: Preschoolers prefer nondissenters as informants. *Psychological Science, 20,* 372–377. http://dx.doi.org/10.1111/j.1467-9280.2009.02291.x

Cosmides, L., & Tooby, J. (2000). Consider the source: The evolution of adaptations for decoupling and metarepresentation. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective* (pp. 53–115). New York, NY: Oxford University Press. Retrieved from https://global.oup.com/academic/product/metarepresentations-9780195141153?cc=us&lang=en&#

Ehrenberg, K., & Klauer, K. C. (2005). Flexible use of source information: Processing components of the inconsistency effect in person memory. *Journal of Experimental Social Psychology, 41,* 369–387. http://dx.doi.org/10.1016/j.jesp.2004.08.001

Erdfelder, E., & Bredenkamp, J. (1998). Recognition of script-typical versus script-atypical information: Effects of cognitive elaboration. *Memory & Cognition, 26,* 922–938. http://dx.doi.org/10.3758/BF03201173

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191. http://dx.doi.org/10.3758/BF03193146

German, T. P., & Barrett, H. C. (2005). Functional fixedness in a technologically sparse culture. *Psychological Science, 16,* 1–5. http://dx.doi.org/10.1111/j.0956-7976.2005.00771.x

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. http://dx.doi.org/10.1146/annurev-psych-122216-011710

Harris, P. L., & Lane, J. D. (2014). Infants understand how testimony works. *Topoi, 33,* 443–458. http://dx.doi.org/10.1007/s11245-013-9180-0

Henrich, J., & McElreath, R. (2003). The evolution of cultural evolution. *Evolutionary Anthropology, 12,* 123–135. http://dx.doi.org/10.1002/evan.10110

Hirshman, E., Whelley, M. M., & Palij, M. (1989). An investigation of paradoxical memory effects. *Journal of Memory and Language, 28,* 594–609. http://dx.doi.org/10.1016/0749-596X(89)90015-6

Hunt, R. R. (1995). The subtlety of distinctiveness: What von Restorff really did. *Psychonomic Bulletin & Review, 2,* 105–112. http://dx.doi.org/10.3758/BF03214414

Inagaki, K., & Hatano, G. (2002). *Young children's naive thinking about the biological world*. New York, NY: Psychology Press. Retrieved from https://www.taylorfrancis.com/books/9780203759844

JASP Team. (2017). JASP (Version 0.9) [Computer software]. Retrieved from https://jasp-stats.org/

Jaswal, V. K., & Neely, L. A. (2006). Adults don't always know best: Preschoolers use past reliability over age when learning new words. *Psychological Science, 17,* 757–758. http://dx.doi.org/10.1111/j.1467-9280.2006.01778.x

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28. http://dx.doi.org/10.1037/0033-2909.114.1.3

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76,* 1261–1277. http://dx.doi.org/10.1111/j.1467-8624.2005.00849.x

Kuhlmann, B. G., Vaterrodt, B., & Bayen, U. J. (2012). Schema bias in source monitoring varies with encoding conditions: Support for a probability-matching account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1365–1376. http://dx.doi.org/10.1037/a0028147

Küppers, V., & Bayen, U. J. (2014). Inconsistency effects in source memory and compensatory schema-consistent guessing. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 67,* 2042–2059. http://dx.doi.org/10.1080/17470218.2014.904914

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4,* 863. http://dx.doi.org/10.3389/fpsyg.2013.00863

Landrum, A. R., Mills, C. M., & Johnston, A. M. (2013). When do children trust the expert? Benevolence information influences children's trust more than expertise. *Developmental Science, 16,* 622–638. http://dx.doi.org/10.1111/desc.12059

Lane, J. D., Wellman, H. M., & Evans, E. M. (2010). Children's understanding of ordinary and extraordinary minds. *Child Development, 81,* 1475–1489. http://dx.doi.org/10.1111/j.1467-8624.2010.01486.x

Larson, H. J., Cooper, L. Z., Eskola, J., Katz, S. L., & Ratzan, S. (2011). Addressing the vaccine confidence gap. *Lancet, 378,* 526–535. http://dx.doi.org/10.1016/S0140-6736(11)60678-8

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., . . . Zittrain, J. L. (2018). The science of fake news. *Science, 359,* 1094–1096. http://dx.doi.org/10.1126/science.aao2998

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind." *Psychological Review, 94,* 412–426. http://dx.doi.org/10.1037/0033-295X.94.4.412

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13,* 106–131. http://dx.doi.org/10.1177/1529100612451018

Mahr, J., & Csibra, G. (2017). Why do we remember? The communicative function of episodic memory. *Behavioral and Brain Sciences, 41,* 1–93. http://dx.doi.org/10.1017/S0140525X17000012

Marsh, R. L., Cook, G. I., & Hicks, J. L. (2006). Gender and orientation stereotypes bias source-monitoring attributions. *Memory, 14,* 148–160. http://dx.doi.org/10.1080/09658210544000015

Mascaro, O., & Sperber, D. (2009). The moral, epistemic, and mindreading components of children's vigilance towards deception. *Cognition, 112,* 367–380. http://dx.doi.org/10.1016/j.cognition.2009.05.012

Mather, M., Johnson, M. K., & De Leonardis, D. M. (1999). Stereotype reliance in source monitoring: Age differences and neuropsychological test correlates. *Cognitive Neuropsychology, 16*(3–5), 437–458. http://dx.doi.org/10.1080/026432999380870

Mayo, R. (2019). Knowledge and distrust may go a long way in the battle with disinformation: Mental processes of spontaneous disbelief. *Current Directions in Psychological Science, 28,* 409–414. http://dx.doi.org/10.1177/0963721419847998

McDaniel, M. A., & Einstein, G. O. (1986). Bizarre imagery as an effective memory aid: The importance of distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 54–65. http://dx.doi.org/10.1037/0278-7393.12.1.54

Mercier, H. (2017). How gullible are we? A review of the evidence from psychology and social science. *Review of General Psychology, 21,* 103–122. http://dx.doi.org/10.1037/gpr0000111

Mercier, H. (2020). *Not born yesterday: The science of who we trust and what we believe*. Princeton, NJ: Princeton University Press. Retrieved from https://press.princeton.edu/books/hardcover/9780691178707/not-born-yesterday

Norenzayan, A., Atran, S., Faulkner, J., & Schaller, M. (2006). Memory and mystery: The cultural selection of minimally counterintuitive narratives. *Cognitive Science, 30,* 531–553. http://dx.doi.org/10.1207/s15516709cog0000_68

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308,* 255–258. http://dx.doi.org/10.1126/science.1107621

Poland, G. A., & Spier, R. (2010). Fear, misinformation, and innumerates: How the Wakefield paper, the press, and advocacy groups damaged the public health. *Vaccine, 28,* 2361–2362. http://dx.doi.org/10.1016/j.vaccine.2010.02.052

RStudio Team. (2020). RStudio (Version 3.6.1) [Computer software]. Retrieved from https://rstudio.com/

Shtulman, A. (2017). *Scienceblind: Why our intuitive theories about the world are so often wrong*. New York, NY: Basic Books. Retrieved from https://www.basicbooks.com/titles/andrew-shtulman/scienceblind/9780465053940/

Shtulman, A., & Carey, S. (2007). Improbable or impossible? How children reason about the possibility of extraordinary events. *Child Development, 78,* 1015–1032. http://dx.doi.org/10.1111/j.1467-8624.2007.01047.x

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science, 14,* 29–56. http://dx.doi.org/10.1207/s15516709cog1401_3

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review, 99,* 605–632. http://dx.doi.org/10.1037/0033-295X.99.4.605

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science, 10,* 89–96. http://dx.doi.org/10.1111/j.1467-7687.2007.00569.x

Sperber, D. (1997). Intuitive and reflective beliefs. *Mind & Language, 12,* 67–83.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language, 25,* 359–393. http://dx.doi.org/10.1111/j.1468-0017.2010.01394.x

Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social and social developmental literatures. *Psychological Bulletin, 111,* 42–61. http://dx.doi.org/10.1037/0033-2909.111.1.42

van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges, 1,* 1600008. http://dx.doi.org/10.1002/gch2.201600008

von Restorff, H. (1933). Über die wirkung von bereichsbildungen im spurenfeld. *Psychologische Forschung, 18,* 299–342. http://dx.doi.org/10.1007/BF02409636

Wertz, A. E. (2019). How plants shape the mind. *Trends in Cognitive Sciences, 23,* 528–531. http://dx.doi.org/10.1016/j.tics.2019.04.009

Wertz, A. E., & Wynn, K. (2014). Selective social learning of plant edibility in 6- and 18-month-old infants. *Psychological Science, 25,* 874–882. http://dx.doi.org/10.1177/0956797613516145

Received December 20, 2019
Revision received August 17, 2020
Accepted August 19, 2020 ■