UNIVERSITY OF CALIFORNIA

Santa Barbara

The mind's meta-data: Cognitive mechanisms for monitoring the source and content of communication

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Psychological & Brain Sciences

by

Spencer J. Mermelstein

Committee in charge:

Professor Tamsin C. German, Chair

Professor Leda Cosmides

Professor Zoe Liberman

Professor Michael B. Miller

December 2021

The dissertation of Spencer J. Mermelstein is approved.

_____

Dr. Leda Cosmides

_____

Dr. Zoe Liberman

_____

Dr. Michael B. Miller

_____

Dr. Tamsin C. German, Committee Chair

December 2021

The mind’s meta-data: Cognitive mechanisms for monitoring the source and content of communication

Spencer J. Mermelstein

To Phyllis J. Porte,

who always wanted to learn more

ACKNOWLEDGEMENTS

I feel immense gratitude to the following for lending their time, energy, brilliance, and friendship to make this dissertation possible. First are the professors who shaped my intellectual life: Mike Miller, for teaching my first class in psychology and offering me my first RAship; Leda Cosmides and John Tooby, for endless inspiration for what psychology can and should be; Zoe Liberman, for bringing clarity to my often vague ideas; Vanessa Woods, for teaching me how to teach; and my advisor, Tamsin German, who decided I needed many more years of education after my undergrad and – in addition to so much else – showed me the value of including everyone into the research process.

The best part of grad school is working alongside completely brilliant people. I'm very grateful for the continual support from my friends in the Cognition & Development Lab and the Center for Evolutionary Psychology: Michael Barlev, for all the years of mentorship and careful thinking that took our projects to the next level; Erin Horowitz, for welcoming me to the CDL and coaching me through grad school and beyond; Jack Strelich, for all the help with stats and coding, and for being an all-around good guy; and Selina Mixner, for creating an awesome lab culture of mentorship. I couldn't do it without my incredible cohort either: Tadeg Quillien and Sakura Arai, thanks for the dear friendship and creating an electric intellectual environment; Evan Layher, for inspiring me with his community service, and Youngki Hong for keeping things in perspective–all friends for life.

I've only made it this far with the love and support from Mom; Angelica and Sergio; Sam and Jehu; Sammy, Tucker, Chloe, and Cha-Cha; Steph and Pepper. Thank you Maggie, you're so brave and I love you – there's nothing we can't do, together.

Spencer J. Mermelstein

Curriculum Vitae as of December 2021

spencer.mer@gmail.com

Education

Ph.D. Psychological & Brain Sciences *2021*
Specialization: Developmental & Evolutionary Psychology
Department of Psychological & Brain Sciences
University of California, Santa Barbara
Advisor: Tamsin C. German, Ph.D.

B.A. Psychology, Minor in Anthropology, with Distinction & High Honors *2014*
Honors Thesis: *Dual representation of concepts in theology and science*
Department of Psychological & Brain Sciences
University of California, Santa Barbara (UCSB)

Publications

Mermelstein, **S**. & German, T. C. (2021). Counterintuitive pseudoscientific beliefs propagate by exploiting the mind's communication evaluation mechanisms. *Frontiers in Psychology*, *12*, doi: https://doi.org/10.3389/fpsyg.2021.739070

Mermelstein, **S**., Barlev, M., & German, T. C. (2020). She told me about a singing cactus: Enhanced memory for the speakers of counterintuitive versus ordinary concepts. *Journal of Experimental Psychology: General*, *150*(5), 972–982. doi: https://doi.org/10.1037/xge0000987

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*(9), doi: https://doi.org/10.1111/cogs.12784

Barlev, M., **Mermelstein**, **S**., & German, T. C. (2018). Representational co-existence 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*(6), 2330–2338. doi: https://doi.org/10.3758/s13423-017-1421-6

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

Mermelstein, S. (2016). The selfish gene. In T. K. Shackelford & V. A. Weekes-Shackelford (Eds.), *Encyclopedia of Evolutionary Psychological Science* (pp. 1–3). NY: Springer. doi: https://doi.org/10.1007/978-3-319-16999-6_1876-1

- Manuscripts in preparation

Mermelstein, **S**., & Barlev, M. & German, T. C. (2021). *The metarepresentation of misinformation: Specialized mechanisms for linking messages that violate prior beliefs to their speakers.* Preprint: https://osf.io/3agkv/

Mermelstein, **S**., & Barlev, M., Alrifai, A., & German, T. C. (2021). *Cultural inputs modulate but do not revise core intuitions: Representational co-existence in Christian and Islamic God concepts.* Preprint: https://osf.io/g3uv5/

Conference Presentations and Invited Talks

Mermelstein, **S**., Barlev, M., & German, T. C. (2021, June). *The mind's meta-data: Cognitive adaptations for monitoring the source of misleading communication*. Paper presented at the annual Meeting of the Human Behavior and Evolution Society, online.

Alrifai, A., **Mermelstein**, **S**., Barlev, M., & German, T. C. (2021, June). *Representational co-existence in Muslims and Christians: Core intuitions about persons interfere with later acquired God concepts across traditions*. Poster presented at the annual Meeting of the Human Behavior and Evolution Society, online.

Mermelstein, **S**., Barlev, M., Alrifai, A., & German, T. C. (2021, May). *Exposure to anthropomorphism explains individual differences in reasoning about God among Christians and Muslims*. Paper presented at the annual California Workshop on Evolutionary Social Sciences, Fullerton, CA [online].

Mermelstein, **S**., (2021, March). *The psychology of conspiracy theories: A multidisciplinary panel discussion*. Participant in public panel discussion hosted by the Pfau Library at CSUSB, San Bernardino, CA [online].

German, T. C., **Mermelstein**, **S**., Barlev, M., Salem, R., & Alrifai, A., (2021, February). *Do God and Allah make mistakes? Representational co-existence in the conception of deities in Christianity and Islam*. Paper presented at the annual Convention of the Society for Personality and Social Psychology, Austin, TX [online].

Mermelstein, **S**., Barlev, M., & German, T. C. (2020, February). *Incoming transmission! Enhanced recall of the senders, but not the recipients, of communicated information that is inconsistent with prior beliefs*. Data-blitz presented at the Evolutionary Psychology Pre-Conference at the annual Convention of the Society for Personality and Social Psychology, New Orleans, LA.

Mermelstein, **S**., Barlev, M., & German, T. C. (2019, June). *Adaptations for evaluating communication explain the representation and transmission of counterintuitive concepts*. Poster presented at the Max Planck Institute for Human Development's annual Summer Institute on Bounded Rationality, Berlin, Germany.

Mermelstein, **S**., Barlev, M., & German, T. C. (2019, June). *An epistemic vigilance framework for the representation and transmission of counterintuitive concepts*. Paper presented at the annual Meeting of the Human Behavior and Evolution Society, Boston, MA.

Mermelstein, **S**. (2019, May). *The application of Bayesian statistics to psychological science using JASP*. Paper presented at the quarterly UCSB seminar for Quantitative Methods in the Social Sciences, Santa Barbara, CA.

Mermelstein, **S.**, & Barlev, M. (2019, February). *Tell me more about that melting lizard! Counterintuitive concepts trigger information search*. Data-blitz presented at the Evolutionary Psychology Pre-Conference at the annual Convention of the Society for Personality and Social Psychology, Portland, OR.

Strelich, J., Wenzel, H., **Mermelstein**, **S.**, & German, T. C. (2019, February). *Successful replication of two theory-of-mind measures in online samples of adults*. Poster presented at the Evolutionary Psychology Pre-Conference at the annual Convention of the Society for Personality and Social Psychology, Portland, OR.

Mermelstein, **S**., Barlev, M., & German, T. C. (2018, July). *Metarepresentation as an adaptation for epistemic vigilance: Enhanced source memory for minimally counterintuitive concepts*. Paper presented at the annual Meeting of the Human Behavior and Evolution Society, Amsterdam, Netherlands.

Mermelstein, **S**., Barlev, M., & German, T. C. (2018, May). *The role of metarepresentation in cultural transmission: Concepts that violate ontological categories remain associated with their source*. Poster presented at the annual California Workshop on Evolutionary Social Sciences, Santa Barbara, CA.

Mermelstein, **S**., Barlev, M., & German, T. C. (2018, May). *Epistemic vigilance: Enhanced recall for the source of minimally counterintuitive concepts*. Poster presented at the annual Convention of the Association for Psychological Science, San Francisco, CA.

Mermelstein, **S**., Barlev, M., & German, T. C. (2017, May). *Adults' reasoning about the properties of God suggests that intuitive inferences are not subject to belief revision*. Poster presented at the annual Meeting of the Human Behavior and Evolution Society, Boise, ID.

Mermelstein, **S**. (2017, May). *The image of God: Psychological and physical intuitions about people underlie the representation of a supernatural entity*. Paper presented at the annual Mini-Convention of the UCSB Department of Psychological and Brain Sciences, Santa Barbara, CA.

Barlev, M., **Mermelstein**, **S**., & German, T. C. (2016, January). *Despite lifelong practice, acquired beliefs about God do not replace conflicting core intuitions about persons*. Poster presented at the Evolutionary Psychology Pre-Conference the annual Convention of the Society for Personality and Social Psychology, San Diego, CA.

Mermelstein, **S**., Barlev, M., & German, T. C. (2015, April). *Despite lifelong practice, reflective religious beliefs do not replace conflicting intuitive inferences in representations of religious concepts*. Poster presented at the annual California Workshop on Evolutionary Social Sciences, San Luis Obispo, CA.

Mermelstein, **S**. (2014, May). *Dual representation of concepts in theology and science*. Poster presented at the annual UCSB Undergraduate Research Colloquium, Santa Barbara, CA.

Teaching Experience

Instructor of Record	*Developmental Psychology*	*2018, 2019, 2021*
	Research Methods	*2019, 2020*
	Department of Psychological & Brain Sciences University of California, Santa Barbara	
Lecturer	*Developmental Psychology*	*2019, 2020*
	Department of Psychology California State University, Channel Islands	
Program Coordinator	*Research Methods* & *Statistics*	*2017-2020*
	Department of Psychological & Brain Sciences University of California, Santa Barbara	

Teaching Assistant	Courses include: *Introduction to STEM for Transfer Students* *Introduction to Psychology* *Research Methods* *Statistics* *Developmental Psychology* *Cognitive Science of Supernatural Concepts* *Lab in Advanced Research Methods* *Lab in Developmental & Evolutionary Psychology* Department of Psychological & Brain Sciences University of California, Santa Barbara	*2015-2021*
Laboratory T.A.	*Research Methods & Experimental Design* Department of Psychology Santa Barbara City College	*2014, 2015*

Mentorship

Undergraduate Honors Thesis advisor for:

Amel Alrifai, B.S. in Psychological & Brain Sciences, UCSB *2021*
Honors Thesis: *Representational co-existence in Islamic God concepts*

Daniel Vu, B.S. in Molecular, Cellular, & Developmental Biology, UCSB *2020*
Honors Thesis: *Students' belief in personal ability buffers the impact of Covid-19 fears on their academic effort regulation*

Hannah Wenzel, B.A. in Psychology, UCSB *2019*
Honors Thesis: *No cognitive benefits for young adult bilinguals*

STEM Coordinator	Summer Teaching Institute for Associates Instructional Development, UCSB	*2020*
Grad Mentor	Teaching Assistant Advisory Program (TAAP) Department of Psychological & Brain Sciences, UCSB	*2018-2021*
Grad Mentor	English for Multilingual Students Program Department of Linguistics, UCSB	*2017-2021*
Undergrad Mentor	AccessGrads: Psychology Mentorship Program Department of Psychological & Brain Sciences, UCSB	*2017-2021*

Awards & Grants

Outstanding Teaching Assistant Award – Nomination, Academic Senate, UCSB – *2020*
Summer Teaching Institute for Associates Certificate, Instructional Development, UCSB – *2019*
Excellence in Teaching Award – Honorable Mention, Grad Student Association, UCSB – *2018*
Doctoral Student Travel Grant, Academic Senate, UCSB – *2018*
Poster Award, California Workshop on Evolutionary Social Science (C-WESS) – *2018*
Graduate Student Association Travel Grant, UCSB – *2017, 2018, 2019*
Distinction in the Psychology Major, UCSB – *2014*
Exceptional Academic Performance Award, UCSB – *2014*
Undergraduate Research and Creative Activities (URCA) Grant, UCSB – *2013*

Professional Activities

Ad-hoc reviewer	*Cognitive Science*	*2020-present*
	Evolutionary Behavioral Sciences	*2020-present*
Invited scholar	Summer Institute on Bounded Rationality	*2019*
	Theme: *Bounded Rationality in a Digital World*	
	Max Planck Institute for Human Development, Berlin, Germany	
Member	*Cognitive Science Society* (CSS)	*2020-present*
	Society for the Teaching of Psychology (STP)	*2019-present*
	Association for Psychological Science (APS)	*2018-present*
	Evolution and Human Behavior Society (HBES)	*2017-present*

Service

Human Subjects Representative *2021*
Committee on Re-opening in Response to COVID-19
Department of Psychological & Brain Sciences, UCSB

ScienceLine Scientist *2016-present*
Designated answerer for questions in the Life Sciences from K-12 students
Materials Research Lab, UCSB

Staff Coordinator *2015*
The Exhibition: *We Remember Them: Acts of Love and Compassion in Isla Vista*
Office of Student Life, UCSB

ABSTRACT

The mind's meta-data: Cognitive mechanisms for monitoring the source and content of communication

by

Spencer J. Mermelstein

Communication is central to our species' success, from facilitating collective action to supercharging cumulative cultural evolution. Yet across human evolutionary history and to the present day, communication carries with it the threat of misinformation and manipulation. For communication to remain adaptive, theorists propose that the mind contains a suite of cognitive mechanisms for evaluating speakers and their messages. This dissertation presents 7 experiments (N = 1,681) investigating a key function hypothesized of these "epistemic vigilance" adaptations: The selective linkage of messages that violate prior beliefs with "meta-data" specifying their source or the context of their acquisition. Remembering such links permits the ongoing evaluation of communication and preserves the integrity of existing knowledge.

In the experiments reported here, participants read a series of stories associated with different sources and each containing counterintuitive (which violate prior beliefs) and ordinary concepts. As predicted, participants in Exp. 1 more accurately attributed counterintuitive versus ordinary concepts to their speakers. Exp. 2a-b replicated this finding

and found that this attribution advantage extended to places and dates associated with counterintuitive concepts. Exp. 3 then investigated the relative strength of these links over time, finding that links between counterintuitive concepts and speakers were differentially durable compared to those with places. Exp. 4 explored the mechanisms that might link epistemically suspect messages to their source. A memory advantage was again found for links between counterintuitive concepts and persons, but only when the messages were framed as told by others ("incoming") and not when told to others ("outgoing"). Exp. 5a-b attempted to replicate this pattern but found an advantage for matching counterintuitive versus ordinary concepts with their associated speakers or recipients, along with an overall advantage for matching incoming versus outgoing messages.

Together, these results demonstrate that the mind selectively tags misleading messages with meta-data that facilitates the ongoing evaluation of their source and content. The results also outline the memory mechanisms – including metarepresentation and elaborative processing – involved in forming links between the source and content of communication. Finally, implications for communication and the representation and social diffusion of counterintuitive concepts broadly, using the case study of those found in pseudoscience, are discussed.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1: INTRODUCTION

Communication yields incredible benefits but also exposes listeners to potentially misguided or misleading information. Indeed, determining what to believe or who to trust is often at the crux of some of the most fateful decisions in life. The net fitness outcomes of these decisions, when played out over phylogenetic time, favored the evolution of cognitive mechanisms for epistemic vigilance. These mechanisms evaluate the source and content of a message to safeguard our knowledge while still allowing for the consideration of information provided by others. This dissertation investigates a key, signature function of the mind's epistemic vigilance mechanisms, the selective linkage of messages that are inconsistent with prior beliefs to their "meta-data" or the context of their acquisition. Chapter 1 sets out the theoretical perspective that informs the current work, puts forward the Source Tagging Hypothesis, and reviews plausible proximate mechanisms. Chapters 2 through 6 detail seven experiments that systematically test predictions generated by the Source Tagging Hypothesis. Chapter 7 discusses the contributions of the current work to the broader study of communication, epistemic vigilance, and counterintuitive concepts, including an application of the theoretical perspective taken here to explain the ubiquity of counterintuitive pseudoscientific concepts across time and cultures.

1.1. Communication and the Cognitive Niche

Humans occupy the cognitive niche, a way of life characterized by the adaptive acquisition, manipulation, and use of vast amounts of information (Barrett, Cosmides, &

Tooby, 2007; Boyd, Richerson, & Henrich, 2011; Cosmides & Tooby, 2000; Tooby & Cosmides, 1992; Tooby & DeVore, 1987). Specifically, and perhaps more than any other species, humans rely upon highly varied, often incomplete, quickly outdated, and socially acquired information that might only be valid within a narrow context. Supporting this mode of life is a set of co-evolved, interlocking cognitive mechanisms that enable humans to deploy inference, social coordination, and language to flexibly leverage the environment to their advantage (Pinker, 2010).

One major source of information about the world comes from communication with other people. A host of evolved and early-developing social learning mechanisms make available the vast quantities of information present in the minds of others. These mechanisms are receptive to the communicative signals of others from infancy (Csibra & Gergely, 2009), facilitate the acquisition of skills that are 'cognitively opaque' to the learner (e.g., tool-making, Csibra & Gergely, 2006), and obviate the need for trial-and-error learning approaches in domains where an error can be catastrophic, such as learning which animals are dangerous (Barrett & Broesch, 2012) or which plants are edible (Wertz & Wynn, 2014). Moreover, communication is the wellspring of cumulative cultural knowledge 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 increasingly abstract scientific and religious concepts (Harris & Koenig, 2006). In short, communication underlies our species success across diverse ecologies and the development of innovations that have changed the world.

The fitness benefits conferred by communication have even sculpted human functional anatomy and brain morphology. Indeed, genomic analyses reveal the rapid

selection of the anatomically modern facial and vocal tract after modern humans split from Denisovans and Neanderthals – features that support speech (Gokhman et al., 2020). Moreover, beginning with Broca, the human brain has been found to contain specialized regions for processing language, with decades of cognitive neuroscience revealing the neural structures and pathways that govern language comprehension and production (Friederici & Gierhan, 2013). Communication thus defines human physical and psychological nature.

1.2. Epistemic Vigilance

A reliance on information from other people, however, creates a vulnerability to being accidently misinformed or intentionally deceived, threatening the existence communication all together. Indeed, evolutionary game theory analyses suggest that populations of credulous communicators may be readily exploited by mendacious mutants (Dawkins & Krebs, 1978; Krebs & Dawkins, 1984; Maynard Smith & Harper, 2003). Communication is therefore always at risk of an evolutionary race to the bottom: Should communication become on average unreliable, listeners may evolve to no longer listen and speakers may evolve to no longer speak.[1] What might keep communication (relatively) reliable rather than rife with misinformation, and therefore beneficial to senders and receivers? While some signals provide evidence of their own reliability, such as the peacock's hard-to-fake tail (Zahavi & Zahavi, 1997), and others are accompanied with credibility-enhancing displays (like eating a mushroom to show others it's not poisonous, Henrich, 2009), these alone may not be sufficient to authenticate the many and varied

[1] This dynamic was pithily summarized by the '80s New Wave group Missing Persons in their song *Words*: "When no one listens / it's no use talkin' at all".

instances of human communication (Mercier & Sperber, 2011). Instead, Sperber and colleagues (2010) propose that human communication is primarily kept advantageous by means of a suite of evolved cognitive mechanisms for *epistemic vigilance* – which function to evaluate the source and content of communication to filter harmful from useful information.

One set of epistemic vigilance mechanisms evaluate the characteristics of a speaker, including their group membership, competency, reliability, and expertise, when judging the veracity of a message. These trust calibration mechanisms have been observed from an early age. Children as young as 5, for instance, adjust their acceptance of messages based on whether an speaker's 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).

Another class of epistemic vigilance mechanisms checks the consistency of message against previous beliefs. Indeed, 4-year-old children have been shown to reject the testimony of a previously reliable informant when their claim (e.g., concerning the location of toy) conflicts with the child's firsthand knowledge (Clément, Koenig, & Harris, 2004). Children also tend to reject claims that conflict with their background knowledge about the properties of objects or animals (Lane & Harris, 2015). Thus, the mind, at even an early age, seems equipped to evaluate the quality of our informants and the fit of their claims against existing beliefs to filter useful from potentially misleading communication.

This research on epistemic vigilance supports the claim that humans are not as gullible as the social sciences have tended to portray (e.g., Gilbert, 1991). For example, Mercier (2017, 2020), in a review of psychological and sociological research, finds that mass influence campaigns, from commercial advertising and political propaganda to religious proselytizing and conspiratorial rumors are rather ineffective in changing people's minds. Nevertheless, it should be noted that epistemic vigilance mechanisms are ecologically rational; that is, they are not objective truth seeking (Mercier, 2017). Our epistemic defenses are potentially vulnerable to evolutionarily novel information ecosystems (e.g., social media), may be overridden by social goals (e.g., as in intergroup conflict), and may even be exploited by psychologically appealing but false beliefs (Mercier, 2017; Mermelstein & German, 2021).

Consequently, the study of epistemic vigilance mechanisms takes on particular urgency in the present day as "fake news," political disinformation, pseudoscientific claims, and conspiracy theories proliferate on online platforms and elsewhere (Lazer et al., 2018). For instance, fraudulent 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 like measles (e.g., Larson, Cooper, Eskola, Katz, & Ratzan, 2011; Poland & Spier, 2010), and misinformation about 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.

1.3. The Source Tagging Hypothesis

This dissertation focuses on the Source Tagging Hypothesis: A key function of epistemic vigilance mechanisms is the selective linkage in memory of messages that violate prior beliefs with meta-data[2] or source tags specifying their speaker or other contextual details such as the place and time such messages were acquired (Cosmides & Tooby, 2000; Mercier, 2017; Johnson, Hashtroudi, & Lindsay, 1993; Sperber, 1997, 2000; Sperber et al., 2010). Source tagging is an instance of source monitoring, defined by Johnson and colleagues (1993) as "expressions of memory that involve judgments about the origin or source of information" with source defined as "a variety of characteristics that specify the conditions under which a memory is acquired (e.g., the spatial, temporal, and social context of the event)".

Tagging communicated information that is inconsistent with existing beliefs with meta-data supports epistemic vigilance in two ways. First, source tags permit the ongoing evaluation of a speaker and the content of their message especially should new information come to light. Second, source tags are critical for demarcating where one's knowledge ends and where the assertions of another begin – thus preserving the integrity of prior beliefs. These two functions of source tags are expanded upon below.

1.3.1. The Ongoing Evaluation of Communication

[2] There's a rich history in psychology, and especially memory research, of comparing mental mechanisms to various technologies of the day, from wax tablets to computers. Of course, these stand as metaphors, but they can still be instructive. Brains really do 'compute' even if how they do so is not like a desktop computer. Meta-data as a technology first became formalized in pre-internet library card catalogs for organizing 'data-about-data,' that is, a book's author, subject, publication date, and so on. Digital meta-data has since become indispensable for organizing and managing databases where information is continually added, drawn-on, and manipulated. Every instance of digital communication, for instance text messages and email, finds itself tagged with meta-data concerning sender, receiver, date of receipt, location, and so on. The point here is that the mind might employ something functionally equivalent to meta-data to aid in the evaluation of communication.

Epistemic vigilance mechanisms are hypothesized to link in memory messages that violate prior beliefs with their speakers and potentially other contextual details. Tracking the speaker of such messages is particularly relevant to epistemic vigilance. By doing so, listeners can continue evaluating these messages in light of new information about their speakers, as well as update their judgements of the competence and trustworthiness of the speakers given new information about the messages. For example, consider a scientist who makes a claim about the dangers of a vaccine that we believe is safe. Our epistemic vigilance mechanisms might specifically link this claim and its speaker for further evaluation. Should, in the future, we find factual errors with their claim, we may then downregulate our judgement of the competence and/or trustworthiness of the scientist and of the accuracy of the claim. On this account, we expect a particularly robust link between speakers and messages that violate pre-existing beliefs, as later acquired information about the speaker or the message might lead us to update our judgments.

The mind's communication evaluation mechanisms may also monitor other contextual details surrounding the acquisition of messages that conflict with our pre-existing beliefs (Cosmides & Tooby, 2000; Johnson et al.,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, we might doubt the accuracy of that scientist's claim about a vaccine more so should we learn

they made that statement while on the payroll of a rival company versus before they were paid by that company. Monitoring the links between speakers and other contextual details associated with inconsistent messages is one key function of source tagging that supports the ongoing evaluation of communication.

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.

1.3.2. Preserving Epistemic Integrity

We now turn toward a second function of source tagging: Tracking the origins of an epistemically suspect message helps differentiates that content from pre-existing beliefs and prevents such messages from immediately revising or updating those beliefs. At the same however we can still comprehend messages that are inconsistent with prior beliefs, further evaluate them, and even reason through their implications – all without accepting them as

true. Holding on to meta-data associated with messages that violate prior beliefs thus serves to preserve the integrity of one's knowledge while allowing us to flexibly consider all the information we acquire. Before detailing how this might be so, I will first review some of the existing scholarship on the psychology of belief, specifically on the question of whether understanding a communicated message requires initially believing that proposition to be true.

Mercier (2020), in a recent update and review of theories of epistemic vigilance, has argued that the mind's communication evaluation faculties might be best described as "open vigilance mechanisms". That is, these mechanisms are not only designed to identify and reject false information, but also to determine whether it is appropriate to update or revise existing knowledge given new information. Indeed, it is an interesting situation when someone tells you something that doesn't fit with your prior beliefs. On one hand, this speaker could have new information that we don't possess, such that it would be important to revise our past beliefs. Alternatively, they could be mistaken or even deceptive, such that updating our prior beliefs would be in error. For an obligatorily communicating species like humans, such situations might have been common over ancestral time and might have impacted fitness: Othello doesn't trust Iago at just his word when the latter implies Desdemona has been unfaithful, but he dares not outright reject the proposition either.[3] In other words, discrepant messages that, if true, hold important consequences need to be carefully considered yet not accepted as fact if they turn out to be false. This presents a challenge to the cognitive mechanisms that evaluate communication. How might these

[3] Othello: "By the world / I think my wife be honest and think she is not. / I think that thou [Iago] art just and think thou art not. / I'll have some proof…" (Shakespeare, 1603, act 3, scene iii).

messages be mentally represented such that we can understand them and reason through their implications, all while at least initially withholding belief?

One influential account suggests that we perhaps *do not* have the cognitive capacity to suspend belief in newly acquired information. Inspired by the philosophy of Spinoza, Gilbert and colleagues (Gilbert, 1991; Gilbert, Krull, & Malone, 1990), posit that comprehending a message entails first accepting it into our beliefs about the world as true and only later and with effort may it be re-evaluated as false. In a line of empirical support for this account, Gilbert and colleagues (1990) tasked participants with learning words in the Hopi language (e.g., "a tica is a fox"). After each learning trial, participants were told whether the statement was true or false. On some trials, however, participants had to additionally respond to a tone played during the presentation of the "true" or "false" label as a means of interrupting the encoding of this information. Later, during a recall task, participants were presented with the previously seen statements and were asked to recall whether each one was labeled true or false. Results revealed that participants were no different at recalling which statements were "true" and which were "false" when encoding of the labels was uninterrupted. However, recall accuracy was differentially reduced for "false" versus "true" statements for trials with an interruption: participants who encoded the labels under cognitive load tended to later mis-identify "false" statements as being "true". Conversely, distractions did not lower recall accuracy for statements labeled "true."

Gilbert (1991) interpreted these findings as evidence in support of a model of belief wherein acquired information, upon comprehension, is by default treated as true and only later may be updated to be false. On this account, distracting participants during the presentation of a falsity label impacted their ability to update the default "true" status of the

statement. These data led Gilbert (1991) to suggest that people, generally, may be quite gullible – a view held widely in the social sciences and popular discourse (for review and critique see Mayo, 2019; Mercier, 2017; Mills, 2013).

Yet this "Spinozan procedure" for processing communication would appear to be poor psychological design for the inhabitants of the cognitive niche. Overly gullible phenotypes would readily invite exploitation, such that communication might quickly cease to (on average) contain reliable information. Ultimately, communication would be selected against as it would no longer benefit senders or receivers (Maynard Smith & Harper, 2003). Yet the massive human reliance on communication over human evolutionary history and to present day, together with evidence of young children's skeptical use of testimony (Harris et al., 2018; Mascaro & Sperber, 2009), suggests the mind may have some more efficient means of evaluating communication.

Moreover, later empirical work has identified strict boundary conditions on the effects observed by Gilbert et al. (1990). For example, Hasson, Simmons, and Todorov (2005) noted that the statement stimuli in Gilbert et al. (1990) were devoid of inferential potential. Whether it is true or false in the Hopi language that "a tica is a fox" was likely irrelevant to the study participants – such a claim does not tap into nor hold consequences for other beliefs. As a result, the mind may not hold on to the truth or falsity tags associated with such inconsequential statements. Under these conditions, a truth bias for information received from others may emerge.

However, it seems unlikely that communication usually concerns wholly irrelevant topics. Instead, the truth value (whether true or false) of a message is likely to impact the status of existing beliefs or combine with them to spur new inferences, including about the

speaker. Researchers have found that the mind can indeed encode a message as false and hold on to that tag should the message hold relevance for one's prior beliefs. For instance, in an experimental design similar to that of Gilbert et al. (1990), Richter, Schroeder, and Wöhrmann (2009) found that statements that tap into strong background beliefs (e.g., "this fast food restaurant offers delicacies") and then labeled as true or false could be encoded and later recalled as such even under conditions of cognitive load. Hasson and colleagues (2005) found that novel, negated information (e.g., "John is a liberal" – FALSE) may be encoded and recalled as false should the very negation of the proposition still yield meaningful inferences. Thus, contrary to earlier accounts, these studies suggest that comprehending a message does not entail representing it as true (for review see Mayo, 2019).

What sort of cognitive architecture might permit a proposition to be understood and its consequences considered – all without being treated as a true datum by other psychological mechanisms? Theorists have proposed that messages received from others are one type of information at least initially held within a *metarepresentational* data format (Cosmides & Tooby, 2000; Leslie, 1987; Sperber, 1994a, 1997, 2000; Sperber & Wilson, 1995). Metarepresentation is the mind's capacity to form a representation of a representation by embedding it within a validating context, a set of meta-data (Sperber, 1997). The validating context can be relatively simple, for example, the metarepresentation "suppose [theory 'x'] is true" contains a representation of a representation (theory 'x') within the context of it being true, regardless of its actual status.[4] Metarepresentations may also take on a more elaborate structure when they contain communicated information.

[4] I thank an anonymous reviewer of a manuscript submitted for publication for this example.

Leslie (1987; see also Klein, German, Cosmides, & Gabriel, 2004; Leslie & Thaiss, 1992) has put forward a specific model of the metarepresentation of communicated propositions. Such metarepresentations link together a representation of an agent (e.g., a person named Sam) and an attitude inferred to be held by that person (e.g., 'believes') toward a communicated proposition (e.g., "it's raining outside"). Thus, Sam's utterance "it's raining outside" can be embedded in the mind of a listener within the metarepresentation "Sam + believes + [it's raining outside]".

Leslie (1987) goes on to suggest that a proposition embedded within a metarepresentational format is 'decoupled' from other beliefs we hold about the world. Specifically, a proposition held in a metarepresentation is suspended from its typical input-output relations and its default truth status with regards to other mechanisms (Leslie, 1987). In this way, we can hold the metarepresentation "Sam believes that [it's raining outside]" while simultaneously believing that it is in fact sunny outside. This protects our pre-existing background beliefs from being updated, revised, or otherwise corrupted by unverified communicated information (Cosmides & Tooby, 2000; Sperber, 1997).

Nonetheless, we can continue to make inferences based on the metarepresentation without accepting the embedded proposition as true. For instance, we can predict Sam will leave home with an umbrella without accepting the premise that it is actually raining. Given these properties, metarepresentation has been implicated in a variety of cognitive processes that involve the suspension of belief, including in forming representations of the mental states of others (mentalizing; Leslie, 1987), imagining counterfactual scenarios (e.g., planning for different future contingencies) or works of fiction (Cosmides & Tooby, 2000),

re-interpreting episodic memories (Mahr & Csibra, 2017), and evaluating communication (Cosmides & Tooby, 2000; Mercier, 2017; Sperber, 1997).

With regards to the evaluation of communication, it is theorized that a message, when first acquired from others, is stored within a metarepresentational format as it is being comprehended (Sperber, 1994a, 1997; Sperber & Wilson, 1995; Mercier, 2017). Thus the message is not initially accepted as true. In the course of comprehending a message, relevant existing beliefs are accessed (Sperber & Wilson, 1995). It is at this point coherence checking epistemic vigilance mechanisms may detect inconsistencies between the message and prior beliefs (Mercier, 2017; Sperber et al., 2010). Should no inconsistencies be found, then the message may lose its meta-representational formatting, including its source tags, and it may be incorporated into our set of existing beliefs (Cosmides & Tooby, 2000). As such, communicated information can quickly result in belief revision in many cases.

A message that violates prior beliefs, however, will continue to be quarantined from other beliefs and remain linked to meta-data like its speaker. Such messages would persist as metarepresentations until sufficient corroborating evidence is gathered to justify updating and replacing prior beliefs, with that evidence coming from personal experience (where possible), trust in the source, and/or a reasoned argument (Mercier & Sperber, 2011). Messages inconsistent with prior beliefs and judged to be false may also continue to be held as metarepresentations for the following reasons: (1) it might be important to retain information you consider false if only to predict how someone who holds that belief might behave and (2) holding on to the link to its speaker serves as important information regarding their credibility. As a last note there may also be propositions, such as counterintuitive concepts in science and religion, that we may reflectively endorse and

consider as true and they yet they remain insulated from other beliefs because they cannot be reconciled with incompatible core knowledge intuitions. More on this later in the section on counterintuitive concepts.

We can now summarize the Source Tagging Hypothesis. Communicated messages received from others are held initially in a metarepresentational format. Epistemic vigilance mechanisms then identify messages that are inconsistent with prior beliefs. Such messages remain quarantined as metarepresentations, which neatly supports epistemic vigilance in two ways. First, the message is linked with meta-data specifying its speaker or other contextual details surrounding its acquisition, permitting the ongoing evaluation of both the message's source and content especially should new information become available. Second, metarepresentation allows us to consider and draw inferences from this information without the risk of corrupting existing beliefs. In this dissertation, I test a key empirical prediction derived from the Source Tagging Hypothesis: messages that violate pre-existing beliefs, more so than messages consistent with those beliefs, remain linked in memory to their speaker and potentially other associated contextual details.

1.4. Mechanisms Underlying Source Tagging

Little of past research has bridged the theoretical construct of metarepresentation with longstanding findings in human memory and source monitoring (for an exception see Mahr & Csibra, 2017). In this section, I suggest how it is that metarepresentations, specifically the hypothesized link between a message and its meta-data, might become memorable.

Following Sperber (1994a, 1997), communicated messages are initially stored in a metarepresentational format as they are comprehended. One consequence of this is that messages received from others are typically stored along with meta-data tags specifying their source. These tags, however, soon decay should the message be judged as consistent with prior beliefs or as irrelevant (Cosmides & Tooby, 2000). Messages found to be at odds with prior beliefs, however, are hypothesized to persist in a metarepresentation and so remain linked to their meta-data (Sperber, 1997). Because they cannot readily be reconciled with prior beliefs, messages that are inconsistent with prior beliefs may continue to draw attention and may be subject to processes of elaborative encoding that enhance the memorability of the metarepresentation in which they are encapsulated (Boyer, 2001; Baumard & Boyer, 2013).

Compatible with this claim, an extensive memory literature beginning with von Restorff (1933) suggests that belief- or expectation-violating information differentially recruits attention, and as such undergoes elaborative processing that facilitates its later recall (for a review see Erdfelder & Bredenkamp, 1998).[5] Indeed, memory advantages have been found for 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,

[5] Elaborative processing (i.e. the "attention-elaboration hypothesis", see Bayen et al., 2000; Erdfelder & Bredenkamp, 1998) encompasses variety of pathways by which atypical, unexpected, or counterintuitive stimuli may become memorable. In response to such stimuli, people may potentially generate distinct inferences (Erdfelder & Kroneisen, 2014), or engage in greater relational processing (think about how the item differs from existing beliefs, Howe & Otgaar, 2013), or think about how oneself might engage with the item (Klein, 2012). I am presently agnostic as to which of these accounts or combination of them may be involved in source tagging. What's critical for the current work is that inconsistent information attracts attention and thus memory.

& Schaller, 2006). Thus, a metarepresentation containing a message that violates prior beliefs might be subject to elaborative processing that heightens its memorability, including the link between the message and its meta-data.

However, some of the literature on source memory finds that attributions of messages to their speakers are often not based on a memory for such links alone, but also on available schematic or stereotypic knowledge (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 memory research finds 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. Note, however, that the memorable discrepancy here is not a consequence of the items themselves being inconsistent with past beliefs, but instead from a schematically unusual combination.

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 studies are the first to explore memory for links between propositions that by themselves violate expectations and their associated contexts. In this way we may test a key hypothesized function of epistemic vigilance mechanisms concerning the meta-data (such as the associated speakers, places, and times) that is 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.

1.5. The Current Study

The current experiments tested the hypothesis that the mind selectively monitors the speaker and potentially other contextual details of messages that are inconsistent with pre-existing beliefs. Of the many classes of communicated information likely to trigger

epistemic vigilance mechanisms, counterintuitive concepts were chosen as a test case, as such concepts regardless of context violate reliably developing and universally-held intuitions about objects, animals, and people (so-called core knowledge intuitions). For example, a message about a "person that can walk through walls" triggers epistemic vigilance mechanisms because it conflicts with core knowledge intuitions about the solidity of bodies and physical objects, an intuition that humans universally hold. Such concepts are therefore likely to be quarantined by epistemic vigilance mechanisms and remain linked to their sources over time. I expand on core knowledge and counterintuitive concepts below.

1.5.1. Counterintuitive Concepts

The human conceptual repertoire is founded in part on reliably developing, species-typical core knowledge mechanisms specialized for representing concepts from fitness-relevant 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"), among other domains (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 only on contact (Baillargeon, 2004; Spelke, Breinlinger, Macomber, & Jacobson, 1992). Infants also interpret and predict the behavior of persons in terms of underlying mental states (Baillargeon et al., 2016), and

understand that beliefs are linked to perceptions, and that people can have beliefs that are false (Onishi & Baillargeon, 2005).

However, 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 instance, the theory of evolution by natural selection violates folk biological intuitions about the immutability of animal "essences"; a statue capable of hearing prayers is a human-made artifact to which a psychological property is transferred (thereby violating folk physical intuitions). While such concepts violate core knowledge intuitions, people may still in certain circumstances come to endorse counterintuitive concepts (such as those that feature prominently in the domains of science and religious theology) and may esteem those who transmit them (e.g., scientists and religious figures).

Nonetheless, Barlev and colleagues (Barlev, Mermelstein, & German, 2017, 2018; Barlev, Mermelstein, Cohen, & German, 2019; Mermelstein, Barlev, Alrifai, & German, in prep) recently presented empirical evidence that even counterintuitive concepts that come to be accepted as true cannot be reconciled with conflicting core knowledge intuitions (also see Barrett, 1998; Barrett & Keil, 1996; Shtulman, 2017). Barlev and colleagues investigated the case study of the God concept among Christian religious adherents. The God concept is initially built by co-opting the person "template," a set of core intuitions about the physical, biological, and psychological properties of people. For example, young children view God as 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 participants evaluated

as “true” or “false” statements that were inconsistent or consistent between core knowledge intuitions about persons and acquired Christian theology about God. As an example, the statement “God has beliefs that are true” is true intuitively and theologically, and the statement “All beliefs God has are false” is false intuitively and theologically. Both items were coded as consistent. In contrast, the statement “God has beliefs that are false” is true intuitively of people but false theologically of God, and the statement “All beliefs God has are true” is false intuitively of people but true theologically of God. Both items were coded as inconsistent. As predicted, Barlev and colleagues found behavioral evidence 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): participants were slower and less accurate at verifying inconsistent statements as compared to consistent statements, even in traditions such as Islam that prohibit all human-like depictions of God (Mermelstein et al., in prep).

Counterintuitive concepts are therefore 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.

1.5.2. Predictions

In the 7 experiments presented here, participants read a series of short stories, with each story containing counterintuitive and ordinary concepts, and each story associated with

different speakers (all Experiments), different places (Experiments 2a and 3), different dates (Experiment 2b), or different recipients (Experiments 4 and 5a-b). After a delay, participants were asked to match each concept to its associated context (i.e., a speaker, place, date, or recipient). Given the goal of investigating memory for links between messages that violate prior beliefs and the context of their acquisition, it was critical that each speaker or other contextual detail was presented with an equal number of counterintuitive and ordinary concepts. Without this feature of the experiments, attributions made during the task might be governed not by remembered links between specific concepts and their speakers, for instance, but by general associations formed between some speakers with counterintuitive concepts and other speakers with ordinary concepts.

Per the Source Tagging Hypothesis, Experiment 1 tested the prediction that counterintuitive concepts would be more accurately attributed to their speakers than ordinary concepts. Experiments 2a and 2b replicate this and test whether counterintuitive concepts associated with different contextual details, places (Experiment 2a), and dates (Experiment 2b), also exhibit a counterintuitive versus ordinary concept attribution accuracy advantage.

The Source Tagging Hypothesis further suggests that epistemic vigilance mechanisms might especially monitor the speakers of messages that are inconsistent with prior beliefs versus other contextual details. One way to test this suggestion was to investigate the durability in memory over time of the links between such messages and their meta-data. Therefore Experiment 3 used a repeated attribution test design with a first attribution phase after a 20-minute delay and a second attribution phase after a 48-hour delay, to examine the relative stability of the links between concepts and their associated contextual details. It was 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 another contextual detail, places.[6]

Finally, this dissertation explores the mechanisms that may underlie the linking of communicated messages to their meta-data. To recap, theories of epistemic vigilance suggest that incoming messages – those sent to us by others – are first held as metarepresentations and so tagged to its speaker. Hypothetically, messages that violate prior beliefs, such as those of counterintuitive concepts, then draw additional attention and processes of elaborative encoding (Bayen et al., 2000; Erdfelder & Bredenkamp, 1998) as they cannot be integrated with prior beliefs. In turn, this heightens the memorability of the metarepresentation and strengthening the link between the message and its speaker.

A test of this metarepresentational account comes from comparing 'destination' memory against source memory; that is, memory for the recipient of a message sent by us to others versus memory for who told us a message (Gopie & MacLeod, 2009). As metarepresentations are hypothesized to have a specialized "slot" for the speaker of a message, we may readily form links between incoming messages and their speakers. However, the metarepresentational data structure is not hypothesized to include a link to the recipient of an outgoing message told by us to others. While a message we tell others might

[6] Experiments 1, 2a-b, and 3 were published as Mermelstein, Barlev, and German (2020), © 2020 by American Psychological Association. Adapted with permission. Citation: Mermelstein, S., Barlev, M., & German, T. C. (2020). She told me about a singing cactus: Enhanced memory for the speakers of counterintuitive versus ordinary concepts. *Journal of Experimental Psychology: General*, *150*(5), 972–982 doi: https://doi.org/10.1037/xge0000987

be meta-represented in our own mind, the data structure would only contain a tag to the self as its speaker, and not to whom it was transmitted (Klein et al., 2004).[7]

The metarepresentational account goes on to suggest that links between speakers and incoming messages that are inconsistent with past beliefs may be particularly memorable as such links may undergo elaborative processing. This differential monitoring of incoming, inconsistent information would functionally support epistemic vigilance: It is specifically these messages that carry the potential threat of misinformation and provide an opportunity to evaluate others.

In Experiments 4 and 5a-b, participants read a series of stories containing counterintuitive and ordinary concepts framed as either told by others (incoming messages) or framed as told to others (outgoing messages). After reading the stories, participants were asked to match each concept to the person it was associated with (either the speakers of incoming messages or the recipients of outgoing messages).

Given the hypothesized format of the metarepresentational data structure, it was predicted that incoming messages, and especially those inconsistent with prior beliefs like counterintuitive concepts, would be more accurately matched to their speakers than outgoing messages were to their recipients. Alternatively, following from more general memory accounts that do not implicate metarepresentation, it might be the case that messages like counterintuitive concepts are highly attention-grabbing and memorable, which in turn heightens the memorability of any of their associated contextual information, regardless of whether they were presented as incoming or outgoing messages.

[7] Although it might be useful to track the recipients of one's own communications in some circumstances this would have to be handled by other mechanisms as the metarepresentational formatting cannot do this within its hypothesized structure.

Chapter 2: EXPERIMENT 1

The Source Tagging Hypothesis states that messages which violate preexisting beliefs remain linked in memory to their speakers. Counterintuitive concepts are one class of information inconsistent with the prior beliefs of people broadly. Thus, it was predicted that counterintuitive concepts would be more accurately attributed to their speakers than ordinary concepts.

2.1. Method

Each experiment in this dissertation uses variations of the method detailed in this section.

2.1.1. Participants

A-priori power analyses were computed for all experiments reported here (for Experiments 1-3 see online at https://osf.io/x5k2u/). 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 dissertation were approved by UCSB's IRB (protocol #23-18-0027) and informed consent was obtained from all participants.

2.1.2. 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.

2.1.3. 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, non-living 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 "had soft fur and liked to play with toys" or the counterintuitive descriptor "had brown spots and could 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 Fig. 1 for sample concepts. See Appendix for all concepts.

Noun Pairs	Domain	Counterintuitive Descriptor
Cat / Dog	Animal	that has brown spots and can walk through solid walls
Shrub / Cactus	Plant	that is small in size and likes to sing loudly
Branch / Rock	Object	that feels cold to the touch and can speak in French
Table / Chair	Artifact	that is big and often floats in midair

Noun Pairs	Domain	Ordinary Descriptor
Cat / Dog	Animal	that has soft fur and likes to play with toys
Shrub / Cactus	Plant	that is dark green and is growing next to a stream
Branch / Rock	Object	that is thick and hard and looks shiny in the sunlight
Table / Chair	Artifact	that is firm to the touch and can hold lots of weight

Fig. 1. Example counterintuitive and ordinary concepts. Counterintuitive concepts contain a violation of core intuitions. Concepts were modified from those in Banerjee et al. (2013).

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, it could be verified upon analysis of the results that attribution accuracy was a function of the type of descriptor (counterintuitive or ordinary), rather than a property of particular 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 below for an example of one of the short stories and Appendix 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 on 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 8 in semi-private computer workstations. Qualtrics software was used to administer all experiments. Data were analyzed using R 3.5.1 and JASP 0.9. Qualtrics scripts, data, power analyses, and R code are available at https://osf.io/x5k2u/. 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 seconds (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 make sure 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 minutes 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 minutes. Fig. 2 summarizes this procedure.

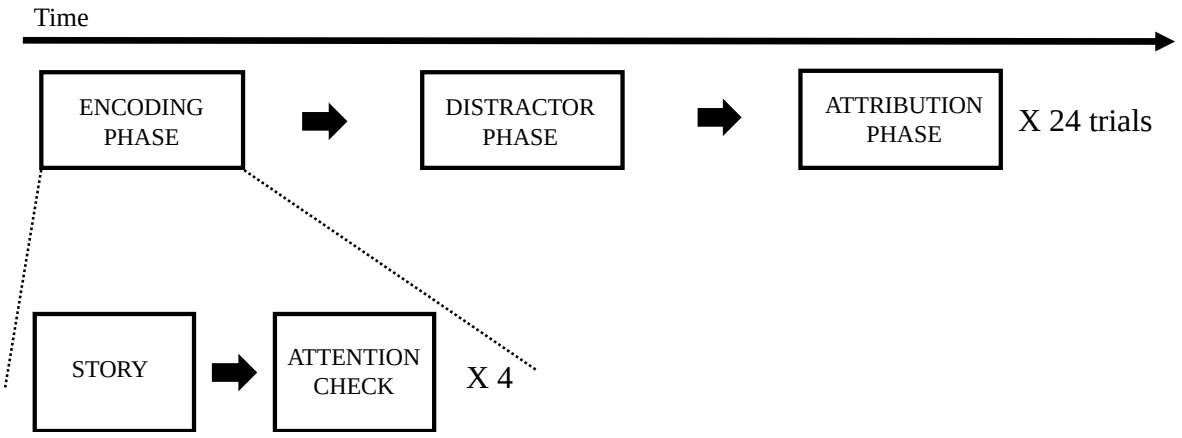

Fig. 2. Summary of the experimental procedure. Participants read four 340-word stories, each containing three counterintuitive and three ordinary concepts, and each associated with a different speaker, other contextual information (places or dates), or recipient. After reading each story, participants in Exp. 1-3 completed an attention check to verify they read and remembered the speaker or other contextual information. In Exp. 1, 2a-b, 4, and 5a there was a distractor phase lasting 2 minutes before the attribution phase, where participants were asked to attribute each concept to the speaker or context with which it was associated. In Exp. 3 there were two attribution phases, one after a distractor phase lasting 20 minutes, and another after a 48 hours delay. Experiment 4 and 5a-b did not have attention checks.

2.2. Results

In this and all other experiments reported in this dissertation 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 Fig. 3 for a pirate plot.

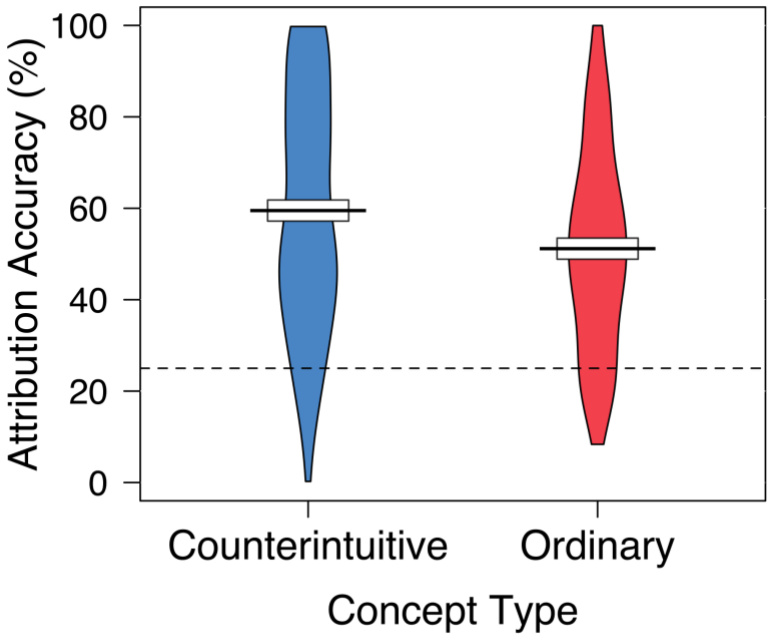

Fig 3. Pirate plot of mean attribution accuracy (%) for counterintuitive and ordinary concepts in Exp. 1. Inference bands correspond to 95% within-subjects CIs. The dotted line at 25% indicates chance performance.

2.3. Discussion

As predicted of the Source Tagging Hypothesis, Exp. 1 found that after a brief delay, counterintuitive concepts were more accurately attributed to their speakers than ordinary concepts. Note that each speaker was associated with an equal number of counterintuitive and ordinary concepts. Thus, participants showed a source attribution advantage not for everything a given person said, but selectively for their messages that were inconsistent with preexisting beliefs. Next, the scope of this memory effect was investigated.

Chapter 3: EXPERIMENT 2a-b

The goal of Exp. 2a-b was to investigate whether epistemic vigilance mechanisms monitor other contextual details, such as the place and time of transmission, associated with messages that violate prior beliefs. Doing so provides a test between two alternative possibilities of what meta-data is linked to messages that violate preexisting beliefs. As argued in the Introduction, 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 to persons. Alternatively, it is nonetheless possible that a broad variety of meta-data remains linked to messages that violate preexisting beliefs (Cosmides & Tooby, 2000; Johnson et al., 1993). 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. Exp. 2a-b tested between these two accounts by comparing the attribution accuracy of counterintuitive versus ordinary concepts linked with speakers versus places (Exp. 2a) and speakers versus dates (Exp. 2b), both after a brief delay.

3.1. Experiment 2a

Experiment 2a compares source memory for speakers versus places associated with the acquisition of counterintuitive and ordinary concepts.

3.1.1. Method

3.1.1.1. Participants

Participants were N = 200 (64% female) UCSB undergraduates (M_{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%).

3.1.1.2. Design

This study used a 2 (Concept: Counterintuitive vs. Ordinary) x 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.

3.1.1.3. Materials and procedure

Participants were randomly assigned to the Person or Place condition. The Person condition was identical to Exp. 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 Exp. 1 except that

participants in the Place condition were asked to attribute each concept to the place where they were told about it.

3.1.2. Results

Attribution accuracy means were entered into a 2 (Concept: Counterintuitive vs. Ordinary) x 2 (Condition: Person vs. Place) mixed ANOVA with repeated measures on the first factor. Results revealed a main effect of Concept [$F(1, 198) = 29.36$, $p < .001$, $\eta^2_p = .13$], 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 to places. See Fig. 4 for pirate plots.

3.2. Experiment 2b

Experiment 2b compares source memory for speakers versus dates associated with the acquisition of counterintuitive and ordinary concepts.

3.2.1. Method

3.2.1.1. Participants

Participants were N = 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%).

3.2.1.2. Design

This study used a 2 (Concept: Counterintuitive vs. Ordinary) x 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.

3.2.1.3. Materials and procedure

Participants were randomly assigned to the Person or Date condition. The Person condition was identical to Exp. 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 Exp. 1 except that participants in the Date condition were asked to attribute each concept to the date on which they were told about it.

3.2.2. Results

Attribution accuracy means were entered into a 2 (Concept: Counterintuitive vs. Ordinary) x 2 (Condition: Person vs. Date) mixed ANOVA with repeated measures on the first factor. Results revealed a main effect of Concept [$F(1, 186) = 24.59$, $p < .001$, $\eta^2_p = .12$], 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 to dates. See Fig. 4 for pirate plots.

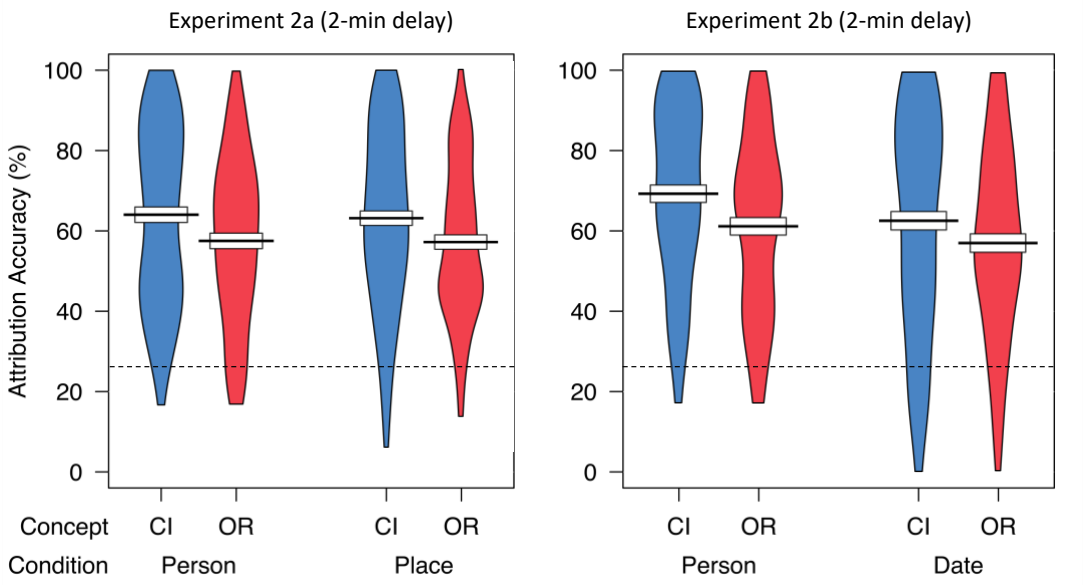

Fig 4. Pirate plots of mean attribution accuracy (%) for counterintuitive (CI) and ordinary (OR) concepts in Exp. 2a and 2b. Inference bands correspond to 95% within-subjects CIs. The dotted line at 25% indicates chance performance.

3.3. Discussion

Exp. 2a-b extended Exp. 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 messages that violate preexisting beliefs. Holding on to this wide set of meta-data (e.g., speakers, places, times) serves epistemic vigilance (Cosmides & Tooby, 2000; Johnson et al., 1993): Recalling both the particular social and spatiotemporal context of an inconsistent message provides important information for the ongoing evaluation of the claim and differentiating that proposition from other beliefs.

Nonetheless, epistemic vigilance mechanisms may be especially likely to hold on to speaker meta-data given their great relevance to the ongoing evaluation of communication. One way to test whether speaker meta-data is preferentially tracked is to investigate the relative durability of links between messages and different varieties of meta-data over time.

Chapter 4: EXPERIMENT 3

Theories of epistemic vigilance suggest that speaker meta-data, all things equal, is more relevant to the ongoing evaluation of communication than other contextual details. To test this, participants in Exp. 3 completed the attribution task twice, once after a short distractor phase (20 minutes) and again after a 48 hours delay. It was 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 to places.

4.1. Method

4.1.1. Participants

Participants were N = 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, n = 194 (92%) returned for the second session. Results from participants who completed both sessions only are reported here. The pattern of results for the first session remains the same if data from the full sample are analyzed.

4.1.2. Design

This study used a 2 (Concept: Counterintuitive vs. Ordinary) x 2 (Delay: 20-minutes vs. 48-hours) x 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.

4.1.3. Materials and procedure

The procedure was identical to that in Exp. 2a, except that after the encoding task, participants completed a 20-minutes (rather than a 2-minutes) battery of distractor tasks before the first attribution task. After 48 hours, 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.

4.2. Results

Attribution accuracy means were entered into a 2 (Concept: Counterintuitive vs. Ordinary) x 2 (Delay: 20-minutes vs. 48-hours) x 2 (Condition: Person vs. Place) mixed ANOVA with repeated measures on the first two factors. Results revealed a main effect of Concept [$F(1, 192) = 24.69$, $p < .001$, $\eta^2_p = 0.11$], a main effect of Delay [$F(1, 192) = 69.39$, $p < .001$, $\eta^2_p = 0.27$], and no main effect of Condition [$F(1, 192) < 1.0$, $p > .250$]. There were no two-way interactions: Concept x Delay [$F(1, 192) < 1.0$, $p > .250$], Condition x Delay [$F(1, 192) < 1.0$, $p > .250$], and Condition x Concept [$F(1, 192) = 3.67$, $p = .057$]. Critically, there was a three-way Concept x Delay x Condition interaction [$F(1, 192) = 10.36$, $p = .002$, $\eta^2_p = 0.05$]. The three-way interaction is unpacked below. See Figure 5 for pirate plots.

Attribution accuracy advantage for persons versus places after 20-minutes and 48-hours. After a 20-minute 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 Exp. 2a. However, after a 48-hour delay, this

attribution accuracy advantage was significantly greater for Persons as compared to 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-minutes, $t(99) = 2.54$, $p = .012$, $d = 0.26$, 95% CI = [0.05, 0.45]; after 48-hours, $t(99) = 5.68$, $p < .001$, $d = 0.57$, 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-minutes, $t(93) = 2.92$, $p = .004$, $d = 0.30$, 95% CI = [0.09, 0.51]; after 48-hours, $t(93) < 1.0$, $p > .250$.

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 ($M_{difference}$ = -5.2%, $SE_{difference}$ = 1.9% vs. $M_{difference}$ = -10.7%, $SE_{difference}$ = 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 ($M_{difference}$ = -5.2%, $SE_{difference}$ = 1.9% vs. $M_{difference}$ = -11.3%, $SE_{difference}$ = 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 ($M_{difference}$ = -11.3%, $SE_{difference}$ = 1.9% vs. $M_{difference}$ = -6.5%, $SE_{difference}$ = 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].

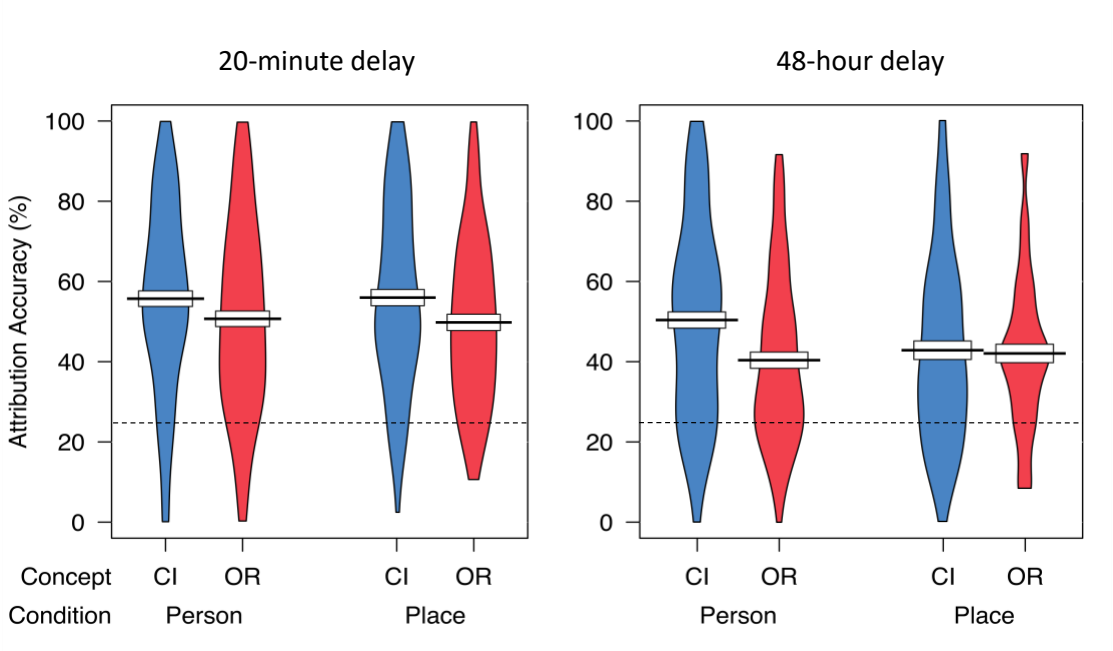

Fig. 5. Pirate plots of mean attribution accuracy (%) for counterintuitive (CI) and ordinary (OR) concepts after 20 minutes and 48 hours in the Person and Place conditions. Inference bands correspond to 95% within-subjects CIs. The dotted line at 25% indicates chance performance.

4.3. Discussion

In sum, after a 20-minute 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-hours, 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 to person-OR links. On the other hand, the counterintuitive versus ordinary concepts attribution accuracy advantage for places disappeared entirely after 48-hours; this was due to a relatively rapid decline of attribution

accuracy over time for place-CI links as compared to place-OR links. These findings are compatible with the suggestion that speaker tags are particularly relevant for the continued evaluation of messages that are inconsistent with prior beliefs.

The preceding 4 experiments provide evidence in support of the Source Tagging Hypothesis: the mind seems to track the context of the acquisition of messages that are inconsistent with prior beliefs, with particularly durable links formed between such messages and their speakers. Next, this dissertation explores the hypothetical mechanisms potentially underlying how the mind tags messages that violate preexisting beliefs with meta-data.

Chapter 5: EXPERIMENT 4

Experiment 4 explores the potential mechanisms underlying the source attribution advantage found for speakers and other contextual details associated with the acquisition of messages inconsistent with prior beliefs (Mermelstein, Barlev, & German, 2020). Theories of epistemic vigilance assert that incoming messages – those sent to us by others – are first quarantined in a metarepresentational formatting (Cosmides & Tooby, 2000; Mercier, 2017; Sperber, 1997). The metarepresentational data structure is proposed to contain specialized "slots" that specify an agent and their attitude toward a proposition (e.g., Sam + believes that + [there is a singing cactus], see Leslie, 1987). Messages found to violate prior beliefs remain sequestered in this format and linked to their speaker, where that link may continually draw attention and recruit memory via elaborative processing (Erdfelder & Bredenkamp, 1998). Quarantining incoming messages in this way functionally supports

epistemic vigilance: It is these messages that may pose a threat to existing knowledge and provide an opportunity to further evaluate our sources.

This metarepresentational account was tested using a memory experiment wherein participants read a series of stories framed either being told by others (incoming messages) or told to others (outgoing messages). Each story contained equal numbers of ordinary and counterintuitive concepts. After reading the stories, participants were asked to match each concept to the person it was associated with (either the speakers of incoming messages or the recipients of outgoing messages).

The metarepresentational data structure is hypothesized to have specialized slots for the speakers of a message said by others but not the recipient of a message we tell others. Thus, this account leads to the unique prediction that counterintuitive concepts will be more accurately matched to their associated person than ordinary concepts, but that this effect will be restricted to cases where the concepts are framed as incoming and not as outgoing messages.

Nonetheless, other memory research suggests alternative, more general mechanisms by which information that violates preexisting beliefs may be linked with associated contextual details. For example, purely associative memory accounts that do not implicate metarepresentation hold that unusual or unexpected information differentially grabs attention, thereby yielding more elaborate processing of this information and encoding in memory of its associated contextual details (Bayen et al., 2000; Ehrenberg & Klauer, 2005; Küppers & Bayen, 2014). This account thus alternatively predicts that counterintuitive concepts would be more accurately matched to their associated persons than ordinary

concepts, regardless of whether they are framed as incoming or outgoing messages. These competing predictions are explored in Experiments 4 and 5a-b.

5.1. Method

5.1.1. Participants

Participants were N = 292 (65% F; M_{age} = 19, SD = 1.27) undergraduates at UCSB who received course credit for their participation. Sample size for 95% power was determined by an a-priori power analysis (available at https://osf.io/3agkv/). Participants identified as East, South, or Southeast Asian (35%), Hispanic (31%), White (26%), or as another ethnic/racial background (8%).

5.1.2. Design

The experiment used a 2 (Concept: Counterintuitive vs. Ordinary) x 2 (Condition: Incoming vs. Outgoing) design with repeated measures on the first factor. The dependent variables were the proportions of counterintuitive and ordinary concepts correctly matched to the person with which they were associated (matching accuracy).

5.1.3. Materials and procedure

Participants were asked to imagine going on a camping trip with several friends: Miguel, Joanna, Sam, and Ariel. Participants were told that "As you're sitting together around the campfire, you and your friends take turns telling each other stories" and to "Try to remember who was taking part in each interaction and what happens in the stories." Participants were randomly assigned to the Incoming or Outgoing condition. In the Incoming condition, each story began with "Your friend [Miguel/Joanna/Sam/Ariel] tells you the following story". This stands as a direct replication of Experiment 1 in this dissertation. In the Outgoing condition, each story began with "You tell your friend [Miguel/Joanna/Sam/Ariel] the following story". The two conditions were otherwise identical.

Participants were tested in groups of up to 8 at computer workstations using Qualtrics software in a testing room at UCSB. The experiment had three phases: encoding, distractor, and test. During the encoding phase, the four stories were displayed one at a time in random order. To encourage careful reading, the computer screen was "locked" on each story for 90 seconds before participants could proceed to the next story. There were no attention checks following each story in this experiment. During the distractor phase, participants typed as many names of the states in the United States as they could for two minutes. Finally, during the test phase, each of the 24 concepts were presented one at a time and in random order. Participants were asked to match each concept to the person with whom it was associated. The experiment took about 20 minutes. All materials, data, and RStudio analysis code for this experiment are available at https://osf.io/3agkv/.

5.2. Results

Concept matching accuracy means were entered into a 2 (Concept: Counterintuitive vs. Ordinary) x 2 (Condition: Incoming vs. Outgoing) mixed ANOVA with repeated measures on the first factor. Results revealed main effects of Concept [$F(1, 290) = 16.01$, $p < .001$, $\eta^2_p = 0.052$] and Condition [$F(1, 290) = 6.47$, $p = .011$, $\eta^2_p = 0.022$], qualified by an interaction between the two [$F(1, 290) = 5.99$, $p = .015$, $\eta^2_p = 0.020$]. See Fig. 6 for a pirate plot.

Simple main effect analyses revealed a matching accuracy advantage for counterintuitive versus ordinary concepts in the Incoming condition, $t(147) = 4.74$, $p < .001$, $d = 0.39$, 95% CI = [0.22, 0.56]. In the Outgoing condition, however, there was no significant difference in matching accuracy between counterintuitive and ordinary concepts, $t(143) = 1.06$, $p = .292$, $d = 0.09$, 95% CI = [-0.08, 0.25].

Finally, planned contrasts revealed that matching accuracy for counterintuitive concepts was significantly higher in the Incoming condition as compared to the Outgoing condition, $t(290) = 3.04$, $p = .003$, $d = 0.36$, 95% CI = [0.13, 0.59]. In contrast, matching accuracy for ordinary concepts did not significantly differ between the Incoming and Outgoing conditions, $t(290) = 1.71$, $p = .088$, $d = 0.20$, 95% CI = [-0.03, 0.43]. Thus, the Concept x Condition interaction was due to differentially accurate memory for the speakers of incoming counterintuitive concepts as compared to the recipients of outgoing counterintuitive concepts or ordinary concepts and their speakers or recipients.

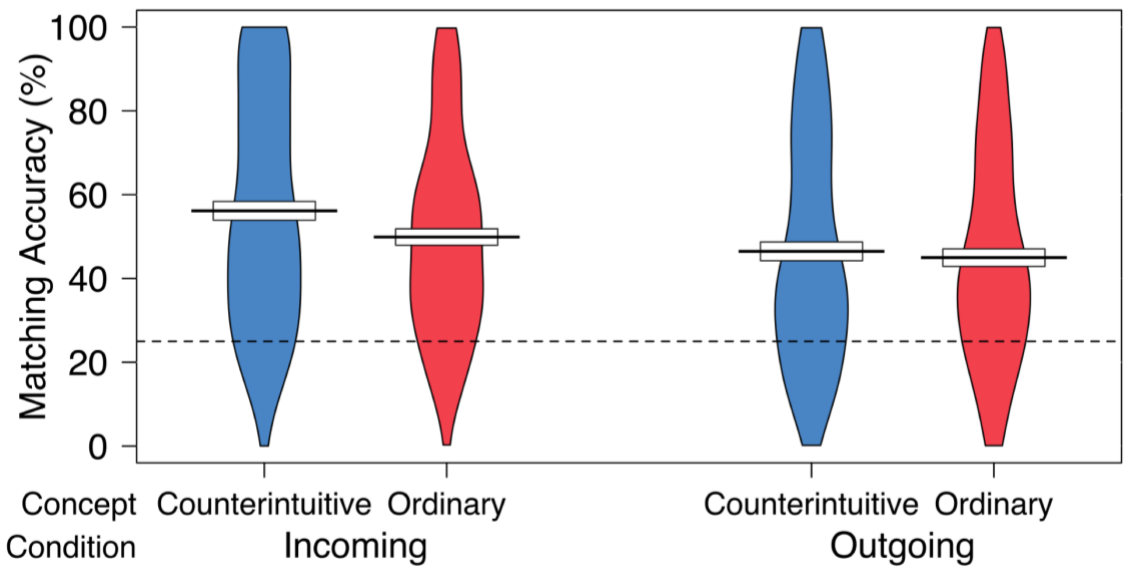

Fig. 6. Pirate plot of mean matching accuracy (%) for Counterintuitive and Ordinary concepts in the Incoming and Outgoing conditions. Inference bands are +/- 1 standard error. The dotted line at 25% indicates chance performance.

5.3. Discussion

Experiment 4 revealed a striking disassociation between minimally different conditions: Participants were more accurate at matching counterintuitive versus ordinary concepts to the persons (speakers or recipients) associated with them, but only when those concepts were framed as incoming as opposed to outgoing messages. Note that these results are not a product of incoming messages in general being better matched to their speaker as ordinary incoming message were no better matched than outgoing messages. Instead, only incoming messages that were inconsistent with prior beliefs exhibited a matching accuracy advantage. Moreover, this pattern discounts general, purely associative memory accounts which would predict no difference in matching accuracy as a function of the incoming versus outgoing framing manipulation.

Instead, this pattern of results is most compatible with a class of theories suggesting that messages that violate preexisting beliefs are linked specifically with their speakers, as predicted of the hypothesized structure of the metarepresentation. Presumably, this linkage may then undergo elaborative processing that heightens its memorability as the message cannot be reconciled with prior beliefs. Moreover, the results strengthen the claim that epistemic vigilance mechanisms use metarepresentation as a part of the evolved cognitive architecture supporting the evaluation of communication. Next, Experiment 5a-b attempts to replicate this pattern of results to increase confidence in the robustness of this metarepresentational account of source tagging.

Chapter 6: EXPERIMENT 5a-b

Experiment 5a-b replicates Experiment 4 with minor design improvements. The primary motivation behind these replications was to increase confidence in the reproducibility of the effects observed in Experiment 4. In another change to the experimental procedure, these studies were conducted remotely, with participants completing the task at a time and place of their choosing because of the Covid-19 pandemic.

6.1. Experiment 5a

Experiment 5a replicates Experiment 4 with a small change to the study's introductory instructions and framing. In Experiment 4, participants were instructed to imagine going camping with a group of friends and that "As you're sitting together around

the campfire, you and your friends take turns telling each other stories." However, during the subsequent encoding phase, each of the four stories displayed was prefaced with a sentence describing a one-on-one interaction rather than a group interaction: in the Incoming condition, "Your friend [Miguel/Joanna/Sam/Ariel] tells you the following story"; in the Outgoing condition, "You tell your friend [Miguel/Joanna/Sam/Ariel] the following story".

It could be possible that this discrepancy between the group interaction framing in the instruction and the one-on-one interaction framing in the encoding phase is possibly a more major concern in the Outgoing condition as compared to the Incoming condition. If so, this discrepancy could perhaps explain, in part, the lower overall accuracy observed in the Outgoing as compared to the Incoming condition: When participants are asked to imagine telling a story to a group (Outgoing condition), it might be strange to then be told that one specific person was the recipient of that story ("You tell your friend [Miguel/Joanna/Sam/Ariel] the following story") as this seems to exclude other recipients. In contrast, when asked to imagine listening to a story in a group (Incoming condition), it is not as strange to imagine being told that you are a recipient of that story ("Your friend [Miguel/Joanna/Sam/Ariel] tells you the following story") as it does not exclude there being other recipients as well.

Experiment 5a corrects for this discrepancy by modifying the instructions so as to set up the expectation for one-on-one interactions during the camping trip. The rest of the study materials are otherwise identical to those of Experiment 4. Furthermore, this study took place after the onset of the global Covid-19 pandemic, such that participants enrolled in psychology courses at UCSB and Arizona State University (ASU) were tested online, at a time and place of their choosing, without proctoring.

6.1.1. Method

6.1.1.1. Participants

The a-priori power analysis used for Experiment 4 was used for Experiments 5a (available at https://osf.io/3agkv/). Based on this power analysis, a minimum of 262 participants (after exclusions, see below) were required to detect an interaction with a small effect size (Cohen's $f = 0.10$) with 95% power and an alpha of 5%.

After exclusions ($n = 27$), participants were $N = 289$ (74% female) UCSB and ASU undergraduates ($M_{age} = 19.7$; $SD = 2.74$), the majority of whom were students at the latter institution ($n = 236$). Participants identified as White (52%), Hispanic or Latino (20%), East, South, or Southeast Asian (19%), or as another ethnic/racial background (9%).

6.1.1.2. Design

As a replication of Experiment 4, Experiment 5a used a 2 (Concept: Counterintuitive vs. Ordinary) x 2 (Condition: Incoming vs. Outgoing) design with repeated measures on the first factor. The dependent variables were the proportions of counterintuitive and ordinary concepts correctly matched to their associated speaker or recipient.

6.1.1.3. Materials and procedure

The experimental materials for Experiment 5a are identical to those in Experiment 4, except for the introductory cover story. Instead of describing stories told around a campfire, Experiment 2 is introduced with the following premise:

> Imagine that you go on a weekend-long camping trip with several close friends: Miguel, Joanna, Ariel, and Sam. When camping, there are many tasks that need to be done: some people get water from the river, others gather wood for the campfire, and yet others boil water for tea or prepare meals. At different times, you and your friends work one on one with each other on these tasks. While working, you tell each other stories to pass the time.

Owing to the COVID-19 pandemic, participants were tested remotely instead of in the lab as in Experiment 1. Despite instructions that emphasized participants pay full attention to the study, the move to an online methodology the likelihood that some participants may be inattentive to the study materials. For this reason, two attention check questions were added to the study, one in the demographics questionnaire and one after the test phase. The questions are “Please select Blue from among the following answer choices: [RED/BLUE/GREEN/YELLOW]" and “What is 20% of 100?”.

Participants’ data were excluded if they failed to provide the correct answer to either of two attention check questions. Second, participants’ data will be excluded should their concept matching accuracy, averaged across counterintuitive and ordinary concepts, be at chance (25%) or worse, as this suggests possible inattentiveness. Exploratory analyses of the results from Experiment 4 revealed that 54 of 292 (18.49%) participants performed on average at chance or worse. Removing these participants did not change the conclusions of Experiment 4; indeed, the key hypothesized 2 x 2 interaction effect size was larger after those exclusions in that study.

6.1.2. Results

Concept matching accuracy means were entered into a 2 (Concept: Counterintuitive vs. Ordinary) x 2 (Condition: Incoming vs. Outgoing) mixed ANOVA with repeated measures on the first factor. Results revealed main effects of Concept, $F(1, 287) = 31.46$, $p < .001$, $\eta^2_p = 0.10$, and Condition, $F(1, 287) = 6.04$, $p = .015$, $\eta^2_p = 0.021$. There was no interaction effect, $F(1, 287) = 1.92$, $p = .167$. Post-hoc tests confirmed that, regardless of condition, Counterintuitive concepts were more accurately matched to their associated persons (speakers or recipients) than Ordinary concepts ($p < .001$). Furthermore, Incoming messages were overall more accurately matched than Outgoing ($p = .015$). See Fig. 7 for a pirate plot.[8]

6.2. Experiment 5b

Experiment 5a failed to replicate the key interaction effect found in Experiment 4. One potential reason was a lack of statistical power as the interaction effect size was smaller than assumed. Experiment 5b again serves as a near direct replication of Experiment 4, but with a larger sample informed by what was observed in Experiment 5a.

[8] As there was no significant interaction in these data simple main effect analyses must be interpreted with caution. Nonetheless, these analyses revealed that the matching accuracy advantage for counterintuitive concepts had a descriptively bigger effect size in the Incoming, $t(136) = 4.90$, $p < .001$, $d = .42$, 95% CI = [0.24, 0.59], versus the Outgoing condition, $t(151) = 3.03$, $p = .003$, $d = .25$, 95% CI = [0.08, 0.41]. While we cannot conclude the effect size was significantly greater in the Incoming versus Outgoing condition and the effect size confidence intervals overlap, this pattern is at least broadly in line with the metarepresentational account. Another observation: Due to participant exclusions, there were 15 fewer participants in the Incoming condition than in the Outgoing condition – such deviations can impact detecting interaction effects, which require as much power as possible (e.g., see https://statmodeling.stat.columbia.edu/2018/03/15/need-16-times-sample-size-estimate-interaction-estimate-main-effect/).

6.2.1. Method

6.2.1.1. Participants

A new power analysis was conducted for Experiment 5b based on the (null) interaction effect size found in Experiment 5a (Cohen's $f = 0.07$). Based on this analysis, a minimum of $N = 400$ participants (after exclusions) were required to detect an interaction with 95% power and an alpha of 5%.

After exclusions ($n = 38$), participants were $N = 393$ (70% female) UCSB and ASU undergraduates ($M_{age} = 20.1$; $SD = 2.87$), the majority of whom were students at the latter institution ($n = 319$). Participants identified as White (43%), Hispanic or Latino (18%), East, South, or Southeast Asian (17%), Black (4%), or as another ethnic/racial background (18%).

6.1.1.2. Design

As a replication of Experiment 4, Experiment 5b used a 2 (Concept: Counterintuitive vs. Ordinary) x 2 (Condition: Incoming vs. Outgoing) design with repeated measures on the first factor. The dependent variables were the proportions of counterintuitive and ordinary concepts correctly matched to their associated speaker or recipient.

6.2.1.3. Materials and procedure

The materials and procedure were the same as in Experiment 5a, except that participants were no longer required to wait 90 seconds before moving on to the next story; they could now move through the stories at their own pace. Participants completed the study at a time and place of their choosing with the instruction that they give the study their full attention.

6.2.2. Results

Concept matching accuracy means were entered into a 2 (Concept: Counterintuitive vs. Ordinary) x 2 (Condition: Incoming vs. Outgoing) mixed ANOVA with repeated measures on the first factor. Results revealed main effects of Concept, $F(1, 391) = 16.19$, $p < .001$, $\eta^2_p = 0.04$, and Condition, $F(1, 391) = 5.25$, $p = .023$, $\eta^2_p = 0.013$. There was no interaction effect, $F(1, 391) = .18$, $p = .671$. Post-hoc tests confirmed that, regardless of condition, Counterintuitive concepts were more accurately matched to their associated persons (speakers or recipients) than Ordinary concepts ($p < .001$). Furthermore, Incoming messages were overall more accurately matched than Outgoing ($p = .023$). Again, the key interaction found in Experiment 4 did not replicate; instead, this pattern of results replicates those found in Experiment 5a. See Fig. 7 for a pirate plot.[9]

[9] Again, simple main effect analyses in this case are suggestive but not definitive. These analyses found a descriptively larger matching accuracy advantage for Incoming, $t(191) = 3.29$ $p = .001$, $d = .24$, 95% CI = [0.09, 0.38], versus Outgoing messages, $t(200) = 2.45$, $p = .015$, $d = .17$, 95% CI = [0.03, 0.31]. However, the effect size confidence intervals do overlap to a great extent, implying no significant difference between these conditions. It is noteworthy, moreover, that the effect sizes observed in this experiment, even in the Incoming condition which serves as a near direct replication of Experiment 1 in this dissertation, were relatively small (e.g., compare $d = .24$ in Exp. 5b-Incoming condition versus $d = .49$ in Exp. 1).

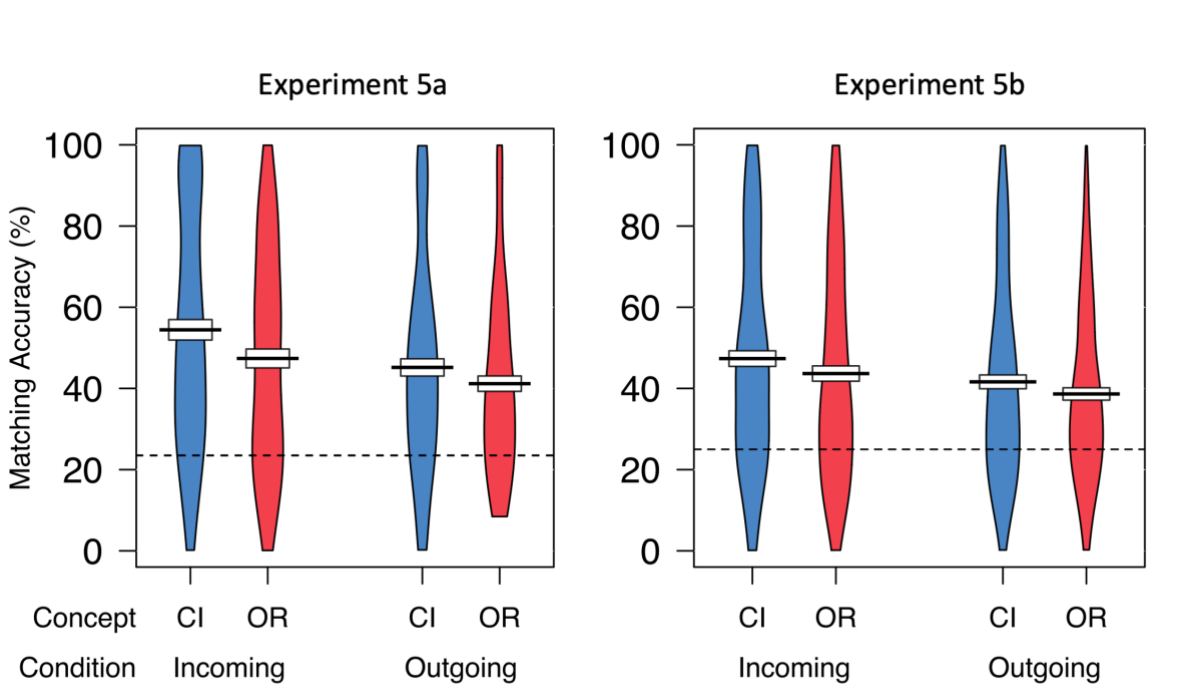

Fig. 7. Pirate plots of mean attribution accuracy (%) for counterintuitive (CI) and ordinary (OR) concepts in the Incoming and Outgoing conditions for Experiments 5a-b. Inference bands are +/- 1 standard error. The dotted line at 25% indicates chance performance.

6.3. Discussion

Experiment 5a-b sought to replicate the pattern of results observed in Experiment 4, which, in a tightly controlled study, found a memory advantage for the speakers of counterintuitive versus ordinary concepts in the Incoming condition, but not for the recipients of those same messages in the Outgoing condition. Such a pattern of results is most compatible with theories suggesting that incoming which violate prior beliefs are

quarantined in a specialized metarepresentational format. Replicating these results would increase confidence in this account.

Methodologically, Experiment 5a-b was the same as Experiment 4 but for (1) minor changes to the introductory text to the study scenario, (2) the participation of primarily ASU undergraduates, and (3) the remote administration of the experiment such that participants completed the task without supervision and on their own time, (4) data exclusions based on attention check failures (these were relatively few), and in the case of Experiment 5b, (5) greater statistical power.

Despite the methodological similarity with Experiment 4, results from Experiment 5a-b failed to replicate the key, hypothesized 2 x 2 interaction effect. Nonetheless, these studies robustly replicated the main effects of Concept, where counterintuitive concepts were overall more accurately matched than ordinary concepts, and Condition, where Incoming messages were overall more accurately matched than Outgoing messages. To better understand the potential methodological issues and theoretical implications at play in these three experiments, the means and standard deviations from Experiments 4 and 5a-b were summarized below in Fig. 8.

	Mean CI (*SD*)		Mean OR (*SD*)	
	Incoming	Outgoing	Incoming	Outgoing
Experiment 4	56% (*27%*)	47% (*27%*)	50% (*24%*)	45% (*25%*)
Experiment 5a	54% (*30%*)	45% (*26%*)	47% (*27%*)	41% (*23%*)
Experiment 5b	47% (*27%*)	42% (*24%*)	44% (*26%*)	39% (*22%*)

Fig. 8. Mean (standard deviation) matching accuracy (%) for counterintuitive (CI) and ordinary (OR) concepts in the Incoming and Outgoing conditions from Experiments 4 and 5a-b.

First, methodological concerns are noted. Fig. 8 reveals an overall decline in concept matching accuracy with each consecutive experiment. One potential exogenous factor explaining this deepening drop in performance may be that Experiment 5a was conducted relatively early in the semester at ASU whereas 5b was conducted late in the semester. The quality of undergraduate data does seem to decline toward the end of the school year, likely made worse in a year of remote instruction and pandemic stressors. Another potential factor for the lowered performance in Experiment 5b was that this study, unlike all others presented here, did not "lock" the stories on screen for a period of time, therefore potentially allowing for participants to speed through the task. This study's attention checks may have attenuated but not eliminated this possibility.

However, the ordering of means from best to worst performance was the same for each study (i.e., Incoming-CI > Incoming-OR > Outgoing-CI > Outgoing-OR). Note that this ordering of means was predicted by the metarepresentational account. While there was no counterintuitive concept matching advantage for the Outgoing condition in Experiment 4, there was one in Experiment 5a-b and of the same magnitude as for the Incoming condition. This was driven by a steeper drop in ordinary concept matching accuracy in the Outgoing versus Incoming conditions from Experiment 4 to 5a. Comparing Experiment 4 and 5b finds that the counterintuitive concept matching accuracy in the Incoming condition experienced the steepest drop, such that the counterintuitive concept matching advantage between the Incoming and Outgoing conditions was of the same magnitude.

The variance in performance in these studies, moreover, is relatively large, suggesting the current method is rather noisy. The drop in performance (perhaps because participants were unsupervised) coupled with the variance inherent to this method was likely

capable of swamping the fragile ordinal interaction (which tend to have relatively small effect sizes) observed in Experiment 4. Best practice for future replications using this method may be to administer this task under supervised laboratory conditions.

Pending additional replications, the modal pattern of results across Experiments 4 and 5a-b require interpretation. Found in each of these studies were main effects of concept type (where CI > OR) and condition (where Incoming > Outgoing). The robustness of these effects holds implications for the mechanisms that underlying source tagging.

Experiment 5a-b provides the first evidence that incoming messages, regardless of fit with prior beliefs, are more accurately matched to their speakers than outgoing messages are to their recipients. This finding is consistent with past research suggesting that incoming information tends to be better integrated in memory with its context than outgoing information (Gopie & MacLeod, 2009; Koriat, Ben-Zur, & Druch, 1991). Gopie and MacLeod (2009) suggested that this is a consequence of where one focuses their attention: When people transmit information to others, their attention tends to be on themselves and the processes involved in generating the information, thus limiting their capacity to associate their message with its recipient. Indeed, manipulations that increased one's self-focus were shown to impair destination memory; on the other hand, manipulations that emphasized the link between a message and a recipient improved destination memory by presumably taking attention off the self (Gopie & MacLeod, 2009).

However, this account raises the question of people may tend to be more self-focused when outputting information in the first place. Speculatively, metarepresentation may provide an explanation: Its data structure is specialized for linking messages to their speakers, but not outgoing messages to their recipients. When the speaker is the self, that

information is applied as the message's source tag. When asked to recall to whom one sent a message, the accessed metarepresentation might only be linked to the self. Remembering outgoing messages thus might be handled by other mechanisms.

On that note, Experiment 5a-b found that, in both the Incoming and Outgoing conditions, counterintuitive concepts were more accurately matched than ordinary concepts and that these effects were no different in magnitude. These results are broadly compatible with accounts of elaborative encoding (e.g., Bayen et al., 2000; Erdfelder & Bredenkamp, 1998). As they are inconsistent with prior beliefs, counterintuitive concepts differentially attract attention and memory, which seems to extend to any associated contextual information such as their speakers or recipients. This finding also demonstrates that manipulations other than emphasizing the identity of the recipient of outgoing information (as in Gopie & MacLeod, 2009) may enhance destination memory. Nonetheless, the effect for Outgoing messages was absent in Experiment 4, and of a descriptively smaller effect in Experiment 5a-b.

Taken together, the results of Experiments 4 and 5a-b highlight the multiple mechanisms likely underlying how the mind monitors communicated information. First, and consistent with accounts of epistemic vigilance, incoming messages seem to undergo greater scrutiny than outgoing messages, with this potentially a signature of the metarepresentation. At the same time, well-studied processes of elaborative memory seem to enhance the memorability of links between messages that are inconsistent with prior beliefs and their associated contextual details broadly.

Chapter 7: GENERAL DISCUSSION

7.1. The Mind's Meta-data

Communication is central to human life: from the coordination of dyadic interactions and multi-person collective action to the social transmission of information about local ecologies and culturally evolved technologies (Boyd, Richerson, & Henrich, 2011; Pinker, 2010). Yet across human evolutionary history and to the present day, communication carries with it the threat of misinformation and manipulation. The impact of political misinformation, "fake news," and conspiracy theories on society stand as glaring modern-day examples (Lazer et al., 2018). Given the threats of misguided or deceptive messages over human evolutionary history, Sperber and colleagues (2010) proposed that humans evolved a suite of epistemic vigilance cognitive mechanisms designed to evaluate messages and their speakers. A key means by which epistemic vigilance mechanisms defend the mind against misinformation is by selectively linking messages that are inconsistent with prior beliefs to their speakers (Mercier, 2017; Sperber, 1997; Sperber et al., 2010).

Linking messages that are inconsistent with prior beliefs with meta-data such as its speaker supports epistemic vigilance in two ways. First, remembering such links enables listeners to update their judgement of a message given new information about its speaker, as well as re-evaluate their opinion of that person, given new information about their message (e.g., Mercier, 2017; Sperber et al., 2010). Second, remembering links between potentially misleading messages and their speakers preserves the integrity of the listener's own beliefs

(e.g., Cosmides & Tooby, 2000; Johnson et al., 1993). In this way, listeners can monitor the origins of such messages and differentiate their own knowledge from the claims of others.

The 7 experiments reported here test the Source Tagging Hypothesis: that we selectively remember the links between messages that are inconsistent with preexisting beliefs and their speakers (Sperber 1997; Sperber et al., 2010; see also Cosmides & Tooby, 2000; Johnson et al., 1993). This hypothesis was tested using the case study of concepts that violate core knowledge intuitions about folk physics, biology, and psychology ("counterintuitive concepts"; e.g., Boyer, 2001). Counterintuitive concepts present an ideal case study because they are inconsistent with the prior beliefs of people broadly (Banerjee et al., 2013; Boyer, 2001; Boyer & Ramble, 2001). Across these 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. After a delay, participants were asked to attribute both concept types to the context of their acquisition.

Experiment 1 found that after a brief delay (2 minutes) participants were more accurate at attributing counterintuitive than ordinary concepts to their speakers. Experiments 2a-b replicated these findings and further found that this attribution accuracy advantage for counterintuitive versus ordinary concepts extended to other contextual details: places (Exp. 2a) and dates (Exp. 2b). Thus, after a brief delay, a broad variety of contextual details were differentially linked in memory to messages that violate preexisting beliefs, which has been suggested to support ongoing monitoring of these messages (e.g., Cosmides & Tooby, 2000; Johnson et al., 1993).

It was predicted, 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. Indeed, the characteristics of a speaker hold great weight on whether we accept what they say (e.g., Harris et al., 2018), such that monitoring them over time may be critical to epistemic vigilance. Given this, we explored the possibility that the links between messages that violate preexisting beliefs and their speakers were especially stable over time compared to links between such messages and other contextual details.

Experiment 3 tested the relative durability of different contextual details associated with counterintuitive concepts using repeated attribution tests. After a short distractor phase (20-minutes), 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-hour 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. Thus, the mind seems particularly prepared to track the speaker of a message that was at odds with preexisting beliefs more so than other associated information, as predicted by theories of epistemic vigilance.

7.1.1. Notes on Mechanism

While Experiments 1-3 empirically establish the Source Tagging Hypothesis, they do not directly investigate the potential cognitive mechanisms responsible for linking messages that are inconsistent with prior beliefs to their contextual details. Theories of communication comprehension assert that acquired messages are first held within a metarepresentational data structure as they are being understood (Sperber, 1994a, 1997; Sperber & Wilson, 1995). As suggested by Leslie (1987), metarepresentation constitutes the minimal cognitive architecture able to decouple representations from one's existing beliefs, thus safeguarding them from erroneous revision. Given this property, metarepresentation has been implicated in many cognitive processes that require representing propositions without accepting them as true, including representations of the mental states of others (Leslie, 1987) and counterfactuals (Cosmides & Tooby, 2000).

Communicated messages, as they are being understood, are hypothesized to first be held as metarepresentations and so linked to meta-data including a link specialized for representing the speaker of the message (Sperber, 1994a, 1997; see also Cosmides & Tooby, 2000; Leslie, 1987; Klein et al., 2004). Messages found to be consistent with past beliefs or supported by other evidence (e.g., from the senses) quickly lose their metarepresentational formatting as the proposition enters into our database of beliefs. On the other hand, messages found to be inconsistent with past beliefs remain quarantined as a metarepresentations. As such messages cannot be readily reconciled with existing beliefs, they continually draw our attention as we (1) look for relevant information to corroborate or challenge the claim and as we (2) adjust our valuation of its speaker. Consequently, a metarepresentation containing an inconsistent message might also be subject to elaborative

processing (Bayen et al., 2000; Erdfelder & Bredenkamp, 1998), heightening the memorability of the whole data structure, including the link between source and content.

Results from Experiments 1-3 are compatible with such a metarepresentation-elaborative processing account. Messages inconsistent with prior beliefs remain were found to remain linked with their meta-data, with especially durable links formed for their speakers. The metarepresentation account, however, makes the additional hypothesis of a specialized link for specifically the speakers of such messages. A comparison of source versus destination memory, that is, memory for messages you received from someone versus a message you send to another, provided a wedge into studying the mechanisms underlying source tagging.

Indeed, if inconsistent messages remain held as metarepresentations, then they should be more accurately attributed to their speakers when received from others than when sent to others. This is because the metarepresentational data structure has a specialized slot for the speaker of a message, but not its recipient. Alternatively, because they are differentially attention-grabbing, counterintuitive concepts may recruit additional processing, heightening the memorability of these concepts and *any* of their associated contextual information compared to ordinary concepts, regardless if the message is received from another or sent to them. The aim of Experiment 4, then, was to test between these metarepresentational-elaboration and pure associative memory accounts. In a tightly controlled experiment, participants read a series of stories containing ordinary and counterintuitive concepts that were framed as either told to them by others ("incoming") or told by them to others ("outgoing"). After a delay, participants matched each concept with the person with which it was associated, speaker or recipient.

Despite a subtle manipulation, Experiment 4 revealed a contrast in matching accuracy for counterintuitive versus ordinary concepts between the incoming and outgoing conditions. As predicted by the metarepresentational account, participants were differentially accurate at matching counterintuitive versus ordinary concepts to their associated persons, but this was only the case for incoming and not outgoing messages. These results support the claim that epistemic vigilance mechanisms monitor specifically incoming messages and affix them with meta-data specifying their speaker. To further grow confidence in this claim, Experiment 5a-b was designed as a near direct replication study but for the notable fact that participants were not tested under laboratory conditions given pandemic restrictions.

While the key interaction found in Experiment 4 did not replicate, Experiment 5a-b provide robust evidence that (1) incoming messages were more accurately matched than outgoing messages and that (2) counterintuitive concepts were overall more accurately matched than ordinary concepts. It is possible that the absence of laboratory control in Experiment 5a-b contributed to the failure to reproduce the interaction observed in Experiment 4. In support of this claim, the pattern of mean scores was similar across experiments yet overall accuracy dropped with each consecutive experiment, potentially swamping the interaction effect. In Experiment 5a, the difference in performance between outgoing counterintuitive and ordinary concepts grew as accuracy for the latter dropped. In Experiment 5b, the difference in performance between incoming counterintuitive and ordinary concepts decreased as accuracy for the former dropped. Pending replications under controlled conditions, let's assume the two main effects are the findings to interpret with regards to identifying the mechanisms underlying source tagging.

The current finding that incoming messages are better integrated with their contextual details than outgoing messages is surprising from a purely associative memory perspective. On that account, the minimal nature of the incoming versus outgoing message manipulation in these studies should not have an effect on context matching. Nonetheless, these results connect with other research on source and destination memory. For example, Gopie and MacLeod (2009) presented participants with pairs of faces and factual statements. In a destination memory condition, participants were instructed to tell, to actually verbalize, the statements to each face; in a source memory condition, participants simply read the statement ostensibly told to them by another person. A later recall task revealed that participants better remembered face and statement pairings in the source versus destination memory condition. In subsequent experiments, Gopie and MacLeod (2009) demonstrated that manipulations to increase one's self-focus tended to further worsen destination memory whereas manipulations that emphasized the identity of a message's recipient improved destination memory. These authors concluded that when people output information, their attention tends to be focused on themselves and the processes required to generate the information, resulting in poorer context integration as compared to source memory.

Experiments 4 and 5a-b find further support for the source versus destination memory advantage. Interestingly, the current experiments found a source advantage without having participants to speak aloud – participants here were asked just to imagine telling or being told each story. Speculatively, imagining outputting information seems like less potent of a manipulation than speaking those messages out loud. That a robust incoming versus outgoing message effect was found under such a minimal manipulation (operationalized as

just few words transposed between conditions) suggests that the mind processes these two classes of information differently.

Gopie and MacLeod (2009) proposed that increased self-focus when outputting information results in typically worse destination as compared to source memory. But this raises the question of why self-focus would increase in the first place when generating a message. Metarepresentation, I contend, might still provide a piece of the answer. Indeed, the metarepresentational data structure is hypothesized to have a slot for representing the speaker of a message, even if that speaker is the self, and not necessarily the recipient of the message. Such an architectural constraint would explain why source memory, by default, seems to be more accurate than destination memory. Manipulations that further enhance self-focus, then, may simply strengthen the existing link between the self and the outgoing message. It is striking that the same stimuli, when framed as inputted versus outputted information, seem to undergo quite different processing that results in their differential integration with their context. Moreover, these data are not compatible with a purely elaborative memory account, which would not predict an incoming versus outgoing effect.

Experiments 4 and 5a-b also highlight the role of elaborative processing in both source and destination memory. These studies found that counterintuitive concepts were more accurately matched regardless of framing as incoming or outgoing messages. Such a finding is compatible with attention-elaboration accounts (Bayen et al., 2000; Erdfelder & Bredenkamp, 1998): Counterintuitive concepts, because they are inconsistent with existing beliefs, recruit attention and heighten the memorability of any associated contextual information. The counterintuitive concept matching advantage was significantly stronger in the Incoming condition in Experiment 4, and descriptively so in Experiment 5a-b. While

tentative, this patten may suggest that incoming information inconsistent with prior beliefs may be particularly subject to monitoring, as expected of a metarepresentation-elaborative processing mechanism for source tagging. Nonetheless, the current studies are the first to demonstrate that transmitting information that violates expectation can improve one's destination memory. An argument could be made that it might be functional to remember the recipients of epistemically-suspect information we transmit – a interesting question to explore in future research.

Collectively, Experiments 4 and 5a-b outline the mechanisms potentially underlying source tagging. Consistent with first principle predictions stemming from epistemic vigilance, the mind differentially monitors incoming versus outgoing information. While the current results are not definitive, the data are suggestive of the role of metarepresentation in accounting for this incoming information advantage. At the same time, the current experiments are compatible with accounts of elaborative encoding: counterintuitive concepts seem to differentially attract attention and so heighten the memorability of any surrounding contextual details. Thus, source tagging might be a result by a variety of underlying mechanisms including those specialized for monitoring communication and general features of human memory.

7.1.2. Implications for Epistemic Vigilance

The current set of experiments advance our understanding of how epistemic vigilance mechanisms monitor and evaluate communication. Epistemic vigilance mechanisms detect inconsistencies between acquired messages and preexisting beliefs and

selectively link these messages to their meta-data, 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.

Linking messages to meta-data 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, these 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), targeting factors such as an audience’s preexisting beliefs may be a productive

starting point in combating the spread of misinformation. The study of the mind's communication evaluation mechanisms is also central to informing theories of the form and diffusion of beliefs broadly. To illustrate, I next present the case study of counterintuitive concepts found in pseudoscience that have been popular across time and cultures.

7.2. Case Study: The Propagation of Counterintuitive Pseudoscience

The epistemic vigilance perspective taken in this dissertation also holds implications for the representation and social transmission of counterintuitive concepts that are culturally widespread. Indeed, concepts ranging from the theory of evolution by natural selection and the electron probability cloud to those of incorporeal spirits and omniscient gods all contain violations of reliably developing intuitions about the world (Boyer, 2001; Shtulman, 2017). Epistemic vigilance mechanisms might target these counterintuitive concepts with consequences for their representational characteristics and social transmission. In the following analysis, the case study of counterintuitive pseudoscientific beliefs is explored. It is argued that counterintuitive concepts, including such pseudoscientific beliefs, propagate by exploiting the mind's epistemic vigilance mechanisms.[10]

Pseudoscience -- claims that take on the guise of scientific knowledge but lack evidentiary support or theoretical plausibility -- is pervasive. At least 40% of Americans, for example, believe in extra-sensory perception and 25% believe that the position of the stars affects life on Earth (Moore, 2005). Pseudoscience can be harmful. The proliferation of anti-vaccination sentiments undermines public health campaigns (Larson et al., 2011) and

[10]This theoretical piece on pseudoscience was published as Mermelstein & German (2021), *Frontiers in Psychology*.

misinformation about global climate change reduces support for mitigation efforts (van der Linden et al., 2017). Understanding the psychological appeal and social transmission of pseudoscience is therefore critical for informing attempts to reduce the impact and spread of these beliefs.

Blancke, Boudry, and colleagues recently advanced a model accounting for the ubiquity of pseudoscience (Blancke et al., 2017, 2019; Blancke & De Smedt, 2013; Boudry et al., 2015). Drawing on Sperber's (1994, 1996) epidemiological theory of cultural representations, these authors have suggested that many forms of pseudoscience are widespread because they cohere with intuitive ways of thinking. For example, those opposed to vaccination often point to pseudoscientific claims that vaccines might cause autism spectrum disorders or other harm (Poland & Spier, 2010). Mercier and Miton (2015) suggest that vaccines, as they entail injecting (inert) pathogens into the body, tap into disgust intuitions that evolved to protect against exposure to contaminants. Vaccines may then be intuitively viewed as a source of contagion, making anti-vaccination claims centered on harm inherently believable, appealing, and transmissible from mind to mind. Other pseudoscientific beliefs may gain traction by exploiting a variety of cognitive predispositions: Creationism/Intelligent Design is grounded in intuitive teleological reasoning (Blancke et al., 2017; Kelemen, 2016); anti-GMO attitudes are based in essentialist intuitions (Blancke et al., 2015); flat earth beliefs are rooted in naive mental models of a geocentric solar system (Vosniadou, 1994).

Along with pseudoscientific beliefs that might exploit a fit with intuitions, however, are a range of such beliefs that manage to spread despite content that is decidedly counterintuitive. Specifically, these 'counterintuitive pseudoscientific' beliefs violate

evolved and reliably developing core knowledge intuitions. Documented as early as infancy (Spelke & Kinzler, 2007), core knowledge intuitions structure our basic expectations of physical objects and their mechanics (e.g., Spelke, 1990) and of intentional agents and their mental states (Baillargeon et al., 2016), among other ontological domains. Thus, counterintuitive concepts are not merely unusual but rather are defined by their incompatibility with the foundational distinctions the mind makes in parsing the world. People may nonetheless acquire counterintuitive concepts; indeed, they are widespread throughout religious, scientific, and pseudoscientific belief systems (Baumard & Boyer, 2013; Boyer, 2001; Shtulman, 2017).

Astrology is one example of counterintuitive pseudoscience. Cultures as diverse as the Babylonians, Han Dynasty China, and the Maya each developed sophisticated belief systems and mathematics to divine the purported influence of the planets and stars on people's personalities and events on Earth (Boxer, 2020). Moreover, astrology remains widespread today despite its contemporary status as a pseudoscience. This is true despite the fact that a central tenet of astrology, that celestial objects can have an influence on people or events on Earth, violates core "folk physics" intuitions that objects cannot act on each other at a distance (Leslie & Keeble, 1987; Spelke, 1990).

Parapsychology, or *psi*, is a second example of counterintuitive pseudoscience. The belief that psychics, mediums, and clairvoyants have a preternatural ability to read minds, manipulate or view distant objects, or tell the future has ancient roots in cultures around the world (Singh, 2018) and has been the subject of research for over 150 years despite its fundamental disconnect from the sciences (Reber & Alcock, 2020). Again, this is despite the fact that these beliefs violate core "folk psychological" intuitions that a person's beliefs are

constrained by their perceptual capacities: that people are ignorant of events they haven't seen or heard (Baillargeon et al., 2016; Onishi & Baillargeon, 2005).

The ubiquity of pseudoscience that contains such drastically counterintuitive elements is potentially surprising from a cultural epidemiology perspective. One reason follows from the suggestion that the prevalence of a belief in a population may depend in part on its fit with intuitive ways of thinking (Sperber, 1994, 1996). On this account, information that is consistent with intuitions is generally more likely to persist across repeated retellings and become more widespread than counterintuitive information (Griffiths et al., 2008; Kalish et al., 2007; Miton et al., 2015; Morin, 2013).

A second potential obstacle to the spread of counterintuitive content stems from the suggestion that the mind contains a host of mechanisms designed to evaluate and filter communicated information (Sperber et al., 2010; see also Mercier, 2017). One function of these epistemic vigilance mechanisms is to assess the plausibility of a message by checking its consistency with prior beliefs. The rudiments of these consistency-checking mechanisms have been documented as early as infancy (Koenig & Echols, 2003), and by age 4, children have been found to reject the claims of others that conflict with their firsthand experiences (Clément et al., 2004) or background knowledge about objects and animals (Lane & Harris, 2015). Counterintuitive information, then, appears to be at a social transmission and believability disadvantage relative to information consistent with cognitive predispositions (Mercier et al., 2019). What then accounts for the cultural success of pseudosciences like astrology and parapsychology?

Here, a pathway by which counterintuitive pseudoscience may spread and receive broad endorsement is proposed. First it is suggested that these beliefs engage the mind's

communication evaluation mechanisms, which largely restrict their influence on behavior. Nonetheless, counterintuitive pseudoscience, as it cannot be fully reconciled with past beliefs, recruits our attention and memory, and triggers a search for more information that may result in the preferential re-transmission of these ideas. During information-search, endorsement of counterintuitive pseudoscience may be bolstered by support from apparently authoritative sources, reasoned arguments, or the functional outcomes of holding such beliefs. Counterintuitive pseudoscience thus achieves cultural prominence by exploiting the mind's communication evaluation mechanisms but explicit belief in such content may not entail tacit commitment.

7.2.1. Representational Format of Counterintuitive Pseudoscience

While communication that is consistent with prior beliefs may be readily accepted, counterintuitive pseudoscience is a class of content that should be flagged by epistemic vigilance mechanisms as requiring further monitoring. By hypothesis, inconsistencies between counterintuitive content and pre-existing beliefs trigger epistemic vigilance mechanisms to quarantine that content from those beliefs via a meta-representational formatting (Sperber, 1997, 2000; see also Mercier, 2017).

As mentioned before in this dissertation, a metarepresentation is a mental data structure that links a proposition to a set of tags that limit the scope of applicability of the information (Cosmides & Tooby, 2000). These tags may take the form of a link to a particular source (Mermelstein et al., 2020), a propositional attitude like certainty or doubt (Leslie, 1987), or a supporting argument (Mercier & Sperber, 2011). For example, the

proposition “the stars influence events on earth” may be embedded in the metarepresentation “my friends believe that [the stars influence events on earth]”. Encapsulated within contextualizing tags, counterintuitive concepts are prevented from spontaneously updating or interacting with existing beliefs or influencing behavior. Nonetheless, one may still come to explicitly profess belief in counterintuitive concepts, deliberately derive inferences from them, and articulate them to others -- but only upon reflection as they cannot be reconciled with conflicting core intuitions (Sperber, 1997).

Epistemic vigilance mechanisms tend to quarantine, rather than outright reject, counterintuitive pseudoscientific beliefs like astrology and parapsychology for two reasons. First, such messages may often be communicated by friends, family, or other influential people. Epistemic vigilance mechanisms are therefore likely to retain these messages (albeit as metarepresentations), given underlying trust in these sources (Sperber, 1997; Sperber et al., 2010; Harris et al., 2018) and social learning biases that motivate people to adopt the beliefs of the successful or prestigious (Henrich & Gil-White, 2001). Relatedly, should a particular counterintuitive concept be widespread in a community, people might at least outwardly endorse such beliefs given that the social cost of rejecting a belief held by their peers may be greater than epistemic costs of harboring them (Hong & Henrich, 2021). Second, epistemic vigilance mechanisms might retain these concepts to aid in the further evaluation of their source and content over time (Mermelstein et al., 2020). Should we later come across information that supports or challenges a given claim, we can then update our judgement of the veracity of the message and the trustworthiness and/or competence of its speaker. Until corroborating evidence is found, we would expect counterintuitive

pseudoscientific concepts to remain quarantined as reflectively-held metarepresentations, with consequences for their stability and capacity to influence behavior.

As reflectively-held beliefs, the counterintuitive concepts found in some varieties of pseudoscience may be variable in their specific content (Baumard & Boyer, 2013). Whereas intuition-consistent pseudoscience might coalesce around a small set of cognitively appealing claims (e.g., "vaccine ingredients cause harm"), counterintuitive beliefs such as "psychics know the future" may be subject to differing and possibly idiosyncratic interpretations. Compatible with this suggestion, proponents of *psi* have put forward a wide range of different accounts for the underlying mechanisms through which these abilities work (Reber & Alcock, 2020). Some accounts, for instance, reference paranormal forces (e.g., a connection to a spirit world), while others may (erroneously) implicate scientific explanations (e.g., quantum mechanics). Without grounding in intuition, the exact content of counterintuitive pseudoscience may be ad-hoc; moreover, these beliefs may be inconsistent or contradictory even within the same mind, as has been documented among adherents of conspiracy theories (Wood et al., 2012) and religious beliefs (Slone, 2007).

Another proposed signature of reflectively-held beliefs is that they may coexist alongside the intuitions with which they conflict rather than update or replace them (Sperber, 1997). Indeed, representational co-existence has been documented for counterintuitive concepts found in science (Kelemen & Rosset, 2009; Shtulman & Harrington, 2016; Shtulman & Valcarcel, 2012) and religion (Barlev et al., 2017, 2018, 2019; Barrett, 1998; Barrett & Keil, 1996). Research on the God concept, for example, finds that religious believers accurately describe God's counterintuitive properties (e.g., omnipresence, omniscience) when explicitly asked, but nonetheless reason as though God

possessed human-like psychology and physicality when indexed by implicit measures (Barrett, 1998; Barrett & Keil, 1996). Co-existence also raises the possibility of interference between mutually incompatible beliefs. Barlev and colleagues (2017, 2018, 2019) asked religious believers to evaluate a series of statements that were consistent or inconsistent in truth-value between intuitions about persons and later-acquired counterintuitive beliefs about God. Participants were slower and less accurate at evaluating inconsistent versus consistent statements, suggesting that intuitions not only co-exist alongside incompatible beliefs, but also conflict with them. The ongoing tension between core intuitions and counterintuitive concepts suggests that these beliefs, including those found in pseudoscience, may not regularly inform behavior.

An implication of this idea is that counterintuitive pseudoscientific concepts might only be deployed in narrow contexts, giving rise to discrepancies between stated beliefs and everyday behavior (Barrett, 1999; Slone, 2007; Sperber, 1985). While one might state their belief that a psychic can tell the future or even follow their horoscope's recommendations when making decisions, they might do so only upon reflection or when prompted. Commitment to counterintuitive pseudoscientific beliefs might generally be at a reflective and not an intuitive level. Indeed, such beliefs may be largely decoupled from behavior as a function of epistemic vigilance mechanisms. A typical believer in *psi*, for instance, would likely make quite different decisions in their life should they implicitly believe that someone could be watching them at any time; the position of the stars and planets may not be one's initial explanation for another's behavior but a post-hoc rationalization. In contrast, intuition-consistent pseudoscience may have a more direct influence on behavior.

Unencumbered by a metarepresentational formatting, anti-vaccination beliefs, for instance, might fluidly translate to vaccine refusal (Mercier & Miton, 2015).

7.2.2. Memory for Counterintuitive Pseudoscience

The memorability of a concept is one predictor of its cultural success: memorable content, all things equal, is more likely to be reproducible and retain fidelity across retellings. Past research suggests that a subset of counterintuitive pseudoscientific beliefs may be mnemonically optimal. Boyer (1994, 2001, 2003) has argued that concepts which are largely consistent with the expectations afforded to ontological categories such as 'person' or 'object' but for a minimal set of violations of those expectations are particularly attention-grabbing, memorable, and inferentially rich. The concept of a ghost fits this 'minimally counterintuitive' template: despite being deceased and capable of passing through solid objects, ghosts are otherwise conceptualized as persons with beliefs and desires. Such striking violations of expectations draw attention as they cannot be fully incorporated into existing beliefs, yet we may still easily imagine and make inferences about ghosts using our knowledge about people. Together, these features make for a differentially memorable combination compared to fully ordinary concepts. Counterintuitive concepts with many violations of expectation (e.g., "a ghost that knows nothing and could never interact with the world"), however, lose their memorability advantage as they cease to hook into existing knowledge and fail to yield many meaningful inferences.

Boyer's (2001) account has received empirical support from laboratory experiments with adults from across cultures (e.g., Nyhof & Barrett, 2001; Boyer & Ramble, 2001) and

with children (Banerjee et al., 2013). Participants in these studies were asked to recall or retell narratives to others, with results demonstrating a memory advantage for minimally counterintuitive (e.g., "a chair that can float in midair") compared to ordinary (e.g., "a table that can hold a lot of weight") or very counterintuitive concepts (e.g., "a rock that could give birth to a singing teapot"). The memory advantage for minimally counterintuitive concepts has also been found to extend to the contextual details associated with them, such as their speaker (Mermelstein et al., 2020). Furthermore, analyses of cultural materials such as folktales from around the world reveal that narratives containing minimally counterintuitive concepts tend to be more common than other concept types (Burdett et al., 2009; Norenzayan et al., 2006). Thus, the mind's attention and memory mechanisms constrain the range of counterintuitive concepts that are likely to be remembered and suitable for cultural success.

It is likely that popular counterintuitive pseudoscientific beliefs are composed of intuitive content alongside compelling, but limited violations of expectation. The wide range of *psi* abilities, for example, seem to be relatively narrow modifications of the capacities typically assumed of persons: supernatural mind-reading may be an overextension of everyday mentalizing, telekinesis an overextension of the expectation that mental states can have effects on the world by directing behavior. Psychics and the like, however, are otherwise conceptualized as ordinary people. The famous psychic Uri Geller could ostensibly bend spoons with his mind, but he nonetheless possessed a physical body that needed to eat, sleep, and breathe. Astrological belief systems may similarly package together counterintuitive and intuitive elements. While the claimed linkage between people and the position of the stars may violate intuitions of cause and effect, astrology does seem to feed

off the human tendencies to perceive patterns in noise (Whitson & Galinsky, 2008), intuit purpose behind complex natural phenomenon (Kelemen et al., 2009), and stereotype others (Lu et al., 2020).

Future empirical work may investigate whether counterintuitive pseudoscientific content strikes a mnemonic optimum for cultural transmission. Laboratory studies employing serial re-transmission methods could demonstrate that minimally counterintuitive pseudoscientific concepts tend to survive repeated retellings compared to other content. Analyses of cultural materials could map out the degree to which astrologers or psychics draw upon counterintuitive versus intuitively-appealing content in making their claims.

7.2.3. Social Re-transmission of Counterintuitive Pseudoscience

The prevalence of a belief in a population, however, depends not only on its memorability but also on individuals being willing to re-transmit it to others. Recent research suggests that people may share counterintuitive concepts with others in an attempt to gather more information about them. Indeed, as early as infancy, violations of expectation have been shown to trigger not only surprise but also information-seeking behavior: Stahl and Feigenson (2015) found that 11-month-old infants who saw an object involved in an counterintuitive event (e.g., a toy appeared to float in midair) preferentially explored that object and manipulated it in an attempt to learn more about its unusual properties in comparison to an ordinary object (e.g., one that fell when unsupported).

The early developing tendency to seek new information in response to a violation of core knowledge may extend across the lifespan, such that one may be motivated to learn

more about counterintuitive concepts, including those found in pseudoscience, in an attempt to reconcile them with prior beliefs. One mode of information-search is to ask others for their opinion, thereby re-transmitting the concept. Compatible with this account, Mermelstein and colleagues (2019) found that novel counterintuitive statements (e.g., "a cactus that liked to sing") were judged by adults to be less believable than ordinary statements (e.g., "a cat that liked to play with toys"), but also as more interesting, more desirable to learn about, and more likely to be passed along to others, and these variables were all strongly correlated. Thus, as with other epistemically suspect information (e.g., "fake news," see Pennycook & Rand, 2021), one's (lack of) belief in counterintuitive content seems to be orthogonal to a willingness to share it with others. People may repeat counterintuitive pseudoscience to others, regardless of their commitment to these beliefs, to scope out what others think about them.[11]

Nonetheless, Mercier and colleagues (2018) have put forward a complementary account suggesting that people may choose to re-transmit pseudoscientific beliefs so as to appear competent to others. Participants in this study rated a series of pseudoscientific (e.g., "people can learn information, like new languages, while asleep") and factual (e.g., "handwriting doesn't reveal personality traits") statements on their believability, one's willingness to re-transmit them, and on how knowledgeable someone who said that statement would seem. A key analysis found that the extent to which a participant believed

[11] Interestingly, the philosopher David Hume (1748/2000) suggested that one may repeat a counterintuitive claim (e.g., of a miracle) that they do not necessarily believe in for the purpose of eliciting "surprise and wonder" in others as to gain their attention and respect. This account is compatible with that of the current paper: a motivation for re-transmitting counterintuitive claims may be to provoke others' reactions to that content. Doing so may reveal whether such claims tend to be endorsed by peers. Moreover, should the counterintuitive claim be favorably received by others (perhaps as it signals a shared group membership), one may be encouraged to continually re-share it to earn their esteem.

that holding a given claim (pseudoscientific or factual) made them appear knowledgeable was an important predictor of their willingness to re-transmit it.

Mercier and colleagues (2018), however, did not differentiate intuition-consistent from counterintuitive pseudoscience. It may be the case that the motive for and method of re-transmission differs depending on the consistency of a claim with core intuitions. Thus, one may be willing to share, and desire to be associated with, intuition-consistent pseudoscience given that others might find that information intuitively compelling (Altay, Claidière, & Mercier, 2020). On the other hand, when re-sharing counterintuitive pseudoscientific beliefs one might tend to attribute them to a source other than the self while gauging others' reactions to that content (Altay, Majima, & Mercier, 2020). For example, disclaimers such as "I read somewhere that..." or "many other people have said..." allow one to discuss counterintuitive ideas with others without asserting ownership of them -- all while promoting the circulation of counterintuitive pseudoscience from mind to mind.

7.2.4. Endoresment of Counterintuitive Pseudoscience

In this analysis, we have distinguished intuition-consistent from counterintuitive pseudoscience, described how the mind's communication evaluation mechanisms might shape the representational characteristics of these counterintuitive concepts, and suggested how such beliefs may become memorable and socially transmissible. We speculate that, as people attempt to reconcile counterintuitive content with their prior beliefs, they may come across different lines of support that lead them to (at least explicitly) accept and endorse counterintuitive pseudoscience.

Mercier and Sperber (2011) have identified two ways by which communication that violates prior beliefs may overcome epistemic vigilance. First, one may suspend their disbelief in such information should they find its source(s) sufficiently trustworthy or reliable. Indeed, scientists have often come to counterintuitive conclusions (e.g., the sun is at the center of the solar system) and laypersons typically trust such claims based on the past reliability and esteem of science in general (Shtulman, 2013). Blancke and colleagues (2019) note that pseudoscience may become believable as it adopts the appearance of science and consequently its privileged epistemic status. Thus, when researchers publish apparent evidence of *psi* in peer reviewed journals, the public may be inclined to believe these claims have a degree of credibility given the source.

Second, acceptance of counterintuitive pseudoscience may come about by encountering supporting argumentation or reasons that justify holding these beliefs (Mercier & Sperber, 2011). Effective arguments in support of counterintuitive pseudoscience might emphasize links between that content and an audience's cognitive predispositions or prior beliefs (Blancke et al., 2019), including existing commitments to other supernatural or paranormal beliefs (Lindeman & Aarnio, 2007). An astrologer's predictions about the future might seem sensible in reference to intuitive pattern-seeking and teleological reasoning tendencies, a psychic might appeal to the widespread and cherished belief in spirits that survive the death of the body in explaining how they communicate with the deceased.

On that note, some strands of counterintuitive pseudoscience may be especially appealing, despite their inconsistency with core intuitions, as they function to alleviate stress or anxiety by providing a compensatory sense of control. Past research finds that many belief systems, from religious beliefs (Inzlicht & Tullet, 2010) and superstitious or magical

thinking (Keinan, 2002) to belief in the efficacy of ritual behavior (Lang et al., 2015), may serve as a buffer against stressful or unpredictable circumstances by offering explanations and actions to take to reduce uncertainty or regain a sense of control (Kay et al., 2009). Interestingly, among pseudosciences, astrology and parapsychology have elements that might serve as anxiolytics. Indeed, experimental work has shown that participants induced to feel that outcomes were out of their control increasingly endorsed the existence of precognition (Greenaway et al., 2013) and followed a psychic's recommendations (Case et al., 2004). Thus, certain counterintuitive pseudoscientific concepts may be particularly likely to gain acceptance, not because their content is intrinsically believable, but because of their functional role in reducing stress or anxiety.

In conclusion, we have argued that counterintuitive pseudoscience has features that exploit the mind's communication evaluation mechanisms to become attention-grabbing, memorable, and likely to be passed on to others. People may even come to explicitly endorse these beliefs through deference to an apparently authoritative source or from a reasoned argument. In this way, counterintuitive pseudoscience achieves cultural prominence. Nonetheless, we hypothesize that these beliefs are held reflectively as they cannot be reconciled with core intuitions (Sperber, 1997). As with other counterintuitive concepts in science (Shtulman & Harrington, 2016; Shtulman & Valcarcel, 2012) and religion (Barlev et al., 2017, 2018, 2019), such pseudoscientific beliefs may coexist alongside incompatible prior beliefs and may be to some extent suspended from guiding behavior. A stated belief in these concepts thus does not necessitate an implicit commitment to them in all contexts. Pseudoscience is ubiquitous but it is not unitary. Recognizing that

these beliefs may propagate through different means may be key to undermining their spread and impact.

7.3. Future Directions

The experiments reported here lay the foundation for future research ranging from detailing the cognitive mechanics of source tagging to examining the functional uses of source tags. First, it would be interesting to further investigate the set of mechanisms underlying the linkage of inconsistent messages to their speakers. Results from Experiments 4 and 5a-b in this dissertation were somewhat mixed as there was a counterintuitive concept matching advantage for outgoing messages in Exp. 5a-b but not in Exp. 4. As seen in Experiment 3, repeated recall tests after longer periods of delay can reveal which linkages are differentially durable in memory. Indeed, the ordering of means across Exp. 4 and 5a-b was constant and as expected by metarepresentational accounts of source tagging such that additional delays might consolidate those trends. Thus, I would propose a replication study of Experiment 4 that makes use of multiple test phases separated by longer delays.

Next, it would be important to ensure that the current findings generalize to other classes of communication information. Counterintuitive concepts were intentionally selected as a test case for the current studies as they should trigger epistemic vigilance mechanisms. Admittedly, counterintuitive concepts, despite their ubiquity in some domains like religion, are an extreme case, almost even a supernormal stimulus for triggering epistemic vigilance mechanisms. Those mechanisms, however, evolved to evaluate more mundane, yet socially relevant messages. Future studies may wish to conceptually replicate the current results

using relatively more ecologically valid stimuli. For example, a more detailed cover story could introduce a social situation such as a workplace dispute. Inconsistencies could be presented between the facts of the scenario and some of the claims of different individuals. The Source Tagging Hypothesis would predict that those messages inconsistent with the facts are likely to remain associated with their speaker. Such a replication might be particularly interesting because the false claims in such a scenario (e.g., someone didn't take the trash out when they said they did) may be attention-grabbing as they clash with reality, but not to the same degree as a counterintuitive concept. More mundane stimuli like this may potentially reduce the degree to which contextual details less relevant to epistemic vigilance get tagged to the statement – thereby revealing the basic design of source tagging mechanisms perhaps obscured by the use of counterintuitive concepts in the current studies.

The functional use of source tags in evaluating communication should also be explored. Future experiments should not only test for source memory of messages the conflict with prior beliefs but also how those memories shape our social interactions with others and trust in their testimony. For example, Klein and colleagues (2009) found evidence that episodic memories serve as boundaries on the extent to which semantic trait judgements about a person are accurate (i.e., Jane tends to be very nice but for those vividly recalled moments when money was involved). Placing source tags on messages that violate preexisting beliefs may similarly guide our future decision as to whom we seek information from. An empirical test of this idea may involve a source memory task with speakers who transmitted varying degrees of incorrect information and then followed by a battery of measures assessing trust in different speakers.

Finally, a key prediction stemming from accounts of epistemic vigilance is that source tags fade when a previously inconsistent message is sufficiently corroborated (Cosmides & Tooby, 2000; Sperber, 1997). It would be very interesting to construct a source memory study similar to those in this dissertation and document a source tagging effect for inconsistent messages, but then randomly assign a group to receive additional information that reconciles the inconsistent information with existing beliefs. At that point, a re-assessment of source memory could reveal that the source tags are then weaker in this group compared to another that did not receive the reconciliatory information.

7.4. Conclusion

Evolutionary cognitive science seeks to understand the evolved design of the human mind – the collection of psychological mechanisms that were selected for over phylogenetic time as a solution to recurrent adaptive problems. How to reap the benefits of communication despite the risk of misguided or manipulative messages stands as one adaptive problem that may have sculpted elements of the mind's functional architecture. Despite their limitations, the current studies demonstrate that theories of cognition grounded in evolutionary logic can lead to testable hypotheses and novel experimental findings. Monitoring the source of messages that violate preexisting beliefs may be crucial to the ongoing evaluation of our social partners and preserving the integrity of our knowledge. Human social life is made possible by communication, and in turn, it is psychological mechanisms like those studied here that ensure communication remains advantageous.

REFERENCES

Altay, S., Claidière, N., & Mercier, H. (2020). It happened to a friend of a friend: Inaccurate source reporting in rumour diffusion. *Evolutionary Human Sciences*, *2*(E49). doi: https://doi.org/10.1017/ehs.2020.53

Altay, S., Majima, Y., & Mercier, H. (2020). It's my idea! Reputation management and idea appropriation. *Evolution & Human Behavior*, *41*(3), 235–243. doi: https://doi.org/10.1016/j.evolhumbehav.2020.03.004

Baillargeon, R. (2004). Infants' physical world. *Current Directions in Psychological Science*, *13*(3), 89–94. doi: https://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. doi: 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 counterintuitive concepts. *Cognitive Science*, *37*, 1251–1289. doi: http://dx.doi.org/10.1111/cogs.12037

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

Barlev, M., Mermelstein, S., & German, T. C. (2018). Representational co-existence 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. doi: https://doi.org/10.3758/s13423-017-1421-6

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*(9). doi: https://doi.org/10.1111/cogs.12784

Barrett, H. C., Cosmides, L., & Tooby, J. (2007). The hominid entry into the cognitive niche. In S. W. Gangestad & J. A. Simpson (Eds.), *Evolution of Mind, Fundamental Questions & Controversies* (pp. 241–248). New York: Guilford Publications.

Barrett, H. C., & Broesch, J. (2012). Prepared social learning about dangerous animals in children. *Evolution & Human Behavior*, *33*(5), 499–508. doi: https://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*(1755), 20122654. doi: https://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*(4), 608–619. doi: https://doi.org/10.2307/1388144

Barrett, J. L. (1999). Theological correctness: Cognitive constraint and the study of religion. *Method & Theory in the Study of Religion*, *11*(4), 325–339. doi: https://doi.org/10.1163/157006899X00078

Barrett, J. L., & Keil, F. C. (1996). Conceptualizing a nonnatural entity: Anthropomorphism in God concepts. *Cognitive Psychology*, *31*(3), 219–247. doi: https://doi.org/10.1006/cogp.1996.0017

Baumard, N., & Boyer, P. (2013). Religious beliefs as reflective elaborations on intuitions: A modified dual-process model. *Current Directions in Psychological Science*, *22*(4), 295–300. doi: https://doi.org/10.1177/0963721413478610

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*(3), 480–500. doi: https://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, & Cognition*, *38*(6), 1512–1529. doi: https://doi.org/10.1037/a0028219

Blancke, S., Boudry, M., & Braeckman, J. (2019). Reasonable irrationality: The role of reasons in the diffusion of pseudoscience. *Journal of Cognition & Culture*, *19*(5), 432–449. doi: https://doi.org/10.1163/15685373-12340068

Blancke, S., Boudry, M., & Pigliucci, M. (2017). Why do irrational beliefs mimic science? The cultural evolution of pseudoscience. *Theoria*, *83*(1), 78–97. doi: https://doi.org/10.1111/theo.12109

Blancke, S., & De Smedt, J. (2013). Evolved to be irrational? Evolutionary and cognitive foundations of pseudosciences. In M. Pigliucci & M. Boudry (Eds.), *The philosophy of pseudoscience* (pp. 361–379). Chicago: The University of Chicago Press.

Blancke, S., Van Breusegem, F., De Jaeger, G., Braeckman, J., & Van Montagu, M. (2015). Fatal attraction: The intuitive appeal of GMO opposition. *Trends in Plant Science*, *20*(7), 414–418. doi: https://doi.org/10.1016/j.tplants.2015.03.011

Boudry, M., Blancke, S., & Pigliucci, M. (2015). What makes weird beliefs thrive? The epidemiology of pseudoscience. *Philosophical Psychology*, *28*(8), 1177–1198. doi: https://doi.org/10.1080/09515089.2014.971946

Boxer, A. (2020). *A scheme of heaven: The history of astrology and the search for our destiny in data*. New York: W. W. Norton & Company.

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, *108*(Supplement 2), 10918-10925. doi: https://doi.org/10.1073/pnas.1100290108

Boyer, P. (1994). *The naturalness of religious ideas: A cognitive theory of religion*. Berkeley, CA: University of California Press.

Boyer, P. (2001). *Religion explained: The evolutionary origins of religious thought*. New York: Basic Books.

Boyer, P. (2003). Religious thought and behaviour as by-products of brain function. *Trends in Cognitive Sciences*, *7*(3), 119–124. doi: https://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. doi: http://dx.doi.org/10.1207/s15516709cog2504_2

Burdett, E. R., Porter, T., & Barrett, J. (2009). Counterintuitiveness in folktales: Finding the cognitive optimum. *Journal of Cognition & Culture*, *9*, 271–287. doi: https://doi.org/10.1163/156770909X12489459066345

Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.

Case, T. I., Fitness, J., Cairns, D. R., & Stevenson, R. J. (2004). Coping with uncertainty: Superstitious strategies and secondary control. *Journal of Applied Social Psychology*, *34*(4), 848–871. doi: https://doi.org/10.1111/j.1559-1816.2004.tb02574.x

Clément, F., Koenig, M., & Harris, P. (2004). The ontogenesis of trust. *Mind & Language*, *19*(4), 360–379. doi: https://doi.org/10.1111/j.0268-1064.2004.00263.x

Cook, J., Ellerton, P., & Kinkead, D. (2018). Deconstructing climate misinformation to identify reasoning errors. *Environmental Research Letters*, *13*, 024018. doi: 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. doi: 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: Oxford University Press.

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Y. Munakata & M. H. Johnson (Eds.), *Processes of Change in Brain and Cognitive Development, Attention & Performance* (pp. 249–274). Oxford: Oxford University Press.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153. doi: https://doi.org/10.1016/j.tics.2009.01.005

Dawkins, R. & Krebs, J. R. (1978). Animal signals: Information or manipulation. In J. R. Krebs & N. B. Davies (Eds.), *Behavioral Ecology: An Evolutionary Approach* (pp. 282–309). Basil Blackwell Scientific.

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. doi: 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*(5), 922–938. doi: http://dx.doi.org/10.3758/BF03201173

Erdfelder, E., & Kroneisen, M. (2014). Proximate cognitive mechanisms underlying the survival processing effect. In B. L. Schwartz, M. Howe, M. Toglia, & H. Otgaar (Eds.), *What is adaptive about adaptive memory?* (pp. 172–198). Oxford University Press.

Friederici, A. D., & Gierhan, S. M. (2013). The language network. *Current Opinion in Neurobiology*, *23*(2), 250-254. doi: https://doi.org/10.1016/j.conb.2012.10.002

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

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, *46*(2), 107–119. doi: https://doi.org/10.1037/0003-066X.46.2.107

Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: Some problems in the rejection of false information. *Journal of Personality & Social Psychology*, *59*(4), 601–613.

Gokhman, D., Nissim-Rafinia, M., Agranat-Tamir, L., Housman, G., García-Pérez, R., Lizano, E., ... & Carmel, L. (2020). Differential DNA methylation of vocal and facial anatomy genes in modern humans. *Nature Communications*, *11*(1), 1-21. doi: https://doi.org/10.1038/s41467-020-15020-6

Gopie, N., & MacLeod, C. M. (2009). Destination memory: Stop me if I've told you this before. *Psychological Science*, *20*(12), 1492–1499. doi: https://doi.org/10.1111/j.1467-9280.2009.02472.x

Greenaway, K. H., Louis, W. R., & Hornsey, M. J. (2013). Loss of control increases belief in precognition and belief in precognition increases control. *PloS ONE*, *8*(8), e71327. doi: https://doi.org/10.1371/journal.pone.0071327

Griffiths, T. L., Kalish, M. L., & Lewandowsky, S. (2008). Theoretical and empirical evidence for the impact of inductive biases on cultural evolution. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*, 3503–3514. doi: http://dx.doi.org/10.1098/rstb.2008.0146

Harris, P. L., & Koenig, M. A. (2006). Trust in testimony: How children learn about science and religion. *Child Development*, *77*(3), 505–524. doi: https://doi.org/10.1111/j.14678624.2006.00886.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. doi: https://doi.org/10.1146/annurev-psych-122216-011710

Hasson, U., Simmons, J. P., & Todorov, A. (2005). Believe it or not: On the possibility of suspending belief. *Psychological Science*, *16*(7), 566–571. doi: https://doi.org/10.1111/j.0956-7976.2005.01576.x

Henrich, J. (2009). The evolution of costly displays, cooperation and religion: Credibility enhancing displays and their implications for cultural evolution. *Evolution & Human Behavior*, *30*(4), 244–260. doi: https://doi.org/10.1016/j.evolhumbehav.2009.03.005

Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution & Human Behavior*, *22*(3), 165–196. doi: https://doi.org/10.1016/S1090-5138(00)00071-4

Henrich, J., & McElreath, R. (2003). The evolution of cultural evolution. *Evolutionary Anthropology*, *12*(3), 123–135. doi: https://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*(5), 594–609. doi: https://doi.org/10.1016/0749-596X(89)90015-6

Hong, Z., & Henrich, J. (2021). The cultural evolution of epistemic practices. *Human Nature*. Online ahead of print. doi: https://doi.org/10.1007/s12110-021-09408-6

Howe, M. L., & Otgaar, H. (2013). Proximate mechanisms and the development of adaptive memory. *Current Directions in Psychological Science*, *22*(1), 16–22. doi: https://doi.org/10.1177/0963721412469397

Hume, D. (1748/2000). *An enquiry concerning human understanding: A critical edition*. Oxford University Press.

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

Inagaki, K., & Hatano, G. (2002). *Young children's naive thinking about the biological world*. New York, NY: Psychology Press.

Inzlicht, M., & Tullett, A. M. (2010). Reflecting on God: Religious primes can reduce neurophysiological response to errors. *Psychological Science*, *21*(8), 1184–1190. doi: https://doi.org/10.1177/0956797610375451

JASP Team (2017). JASP [Computer software].

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. doi: 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. doi: http://dx.doi.org/10.1037/0033-2909.114.1.3

Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin & Review*, *14*, 288–294. doi: http://dx.doi.org/10.3758/BF03194066

Kay A. C., Whitson J. A., Gaucher D., & Galinsky A. D. (2009). Compensatory control: Achieving order through the mind, our institutions, and the heavens. *Current Directions in Psychological Science*, *18*, 264–268. doi: https://doi.org/10.1111/j.1467-8721.2009.01649.x

Keinan, G. (2002). The effects of stress and desire for control on superstitious behavior. *Personality & Social Psychology Bulletin*, *28*(1), 102–108. doi: https://doi.org/10.1177/0146167202281009

Kelemen, D. (2004). Are children "intuitive theists"? Reasoning about purpose and design in nature. *Psychological Science*, *15*(5), 295–301. doi: https://doi.org/10.1111/j.0956-7976.2004.00672.x

Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanation in adults. *Cognition*, *111*(1), 138–143. doi: https://doi.org/10.1016/j.cognition.2009.01.001

Kelemen, D., Rottman, J., & Seston, R. (2012). Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default. *Journal of Experimental Psychology: General*, *142*, 1074–1083. doi: https://doi.org/10.1037/a0030399

Klein, S. B. (2012). A role for self-referential processing in tasks requiring participants to imagine survival on the savannah. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *38*, 1234–1242. doi: https://doi.org/10.1037/a0027636

Klein, S. B., Cosmides, L., Gangi, C. E., Jackson, B., Tooby, J., & Costabile, K. A. (2009). Evolution and episodic memory: An analysis and demonstration of a social function of episodic recollection. *Social Cognition*, *27*(2), 283–319. doi: https://doi.org/10.1521/soco.2009.27.2.283

Klein, S. B., German, T. P., Cosmides, L., & Gabriel, R. (2004). A theory of autobiographical memory: Necessary components and disorders resulting from their loss. *Social Cognition*, *22*(5), 460–490. doi: https://doi.org/10.1521/soco.22.5.460.50765

Koenig, M. A., & Echols, C. H. (2003). Infants' understanding of false labeling events: The referential roles of words and the speakers who use them. *Cognition*, *87*(3), 179–208. doi: https://doi.org/10.1016/S0010-0277(03)00002-7

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*(6), 1261–1277. doi: https://doi.org/10.1111/j.1467-8624.2005.00849.x

Koriat, A., Ben-Zur, H., & Druch, A. (1991). The contextualization of input and output events in memory. *Psychological Research*, *53*(3), 260–270. doi: https://doi.org/10.1007/BF00941396

Krebs, J. R. & Dawkins, R. (1984). Animal signals: Mind-reading and manipulation. In J. R. Krebs & N. B. Davies (Eds.), *Behavioral Ecology: An Evolutionary Approach* (pp. 380–402). Blackwell Science.

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, & Cognition, 38, 1365–1376. doi: http://dx.doi.org/10.1037/a0028147

Küppers, V., & Bayen, U. J. (2014). Inconsistency effects in source memory and compensatory schema-consistent guessing. *The Quarterly Journal of Experimental Psychology, 67*(10), 2042–2059. doi: https://doi.org/10.1080/17470218.2014.904914

Landrum, A. R., Mills, C. M., & Johnston, A. M. (2013). When do children trust the expert? Benevolence information influences children's trust more than expertise. *Developmental Science, 16*(4), 622–638. doi: https://doi.org/10.1111/desc.12059

Lane, J. D., & Harris, P. L. (2015). The roles of intuition and informants' expertise in children's epistemic trust. *Child Development, 86*(3), 919–926. doi: https://doi.org/10.1111/cdev.12324

Lane, J. D., Wellman, H. M., & Evans, E. M. (2010). Children's understanding of ordinary and extraordinary minds. *Child Development, 81*(5), 1475–1489. doi: https://doi.org/10.1111/j.1467-8624.2010.01486.x

Lang, M., Krátký, J., Shaver, J. H., Jerotijević, D., & Xygalatas, D. (2015). Effects of anxiety on spontaneous ritualized behavior. *Current Biology, 25*(14), 1892–1897. doi: https://doi.org/10.1016/j.cub.2015.05.049

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., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., . . . & Schudson, M. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096. doi: https://doi.org/10.1126/science.aao2998

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

Leslie, A. M., & Keeble, S. (1987). Do six-month-old infants perceive causality? *Cognition*, *25*(3), 265–288. doi: https://doi.org/10.1016/S0010-0277(87)80006-9

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition*, 43(3), 225–251. doi: https://doi.org/10.1016/0010-0277(92)90013-8

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*(3), 106–131. doi: https://doi.org/10.1177/1529100612451018

Lindeman, M., & Aarnio, K. (2007). Superstitious, magical, and paranormal beliefs: An integrative model. *Journal of Research in Personality*, *41*(4), 731–744. doi: https://doi.org/10.1016/j.jrp.2006.06.009

Lu, J. G., Liu, X. L., Liao, H., & Wang, L. (2020). Disentangling stereotypes from social reality: Astrological stereotypes and discrimination in China. *Journal of Personality & Social Psychology*, 119(6), 1359–1379. doi: https://doi.org/10.1037/pspi0000237

Mahr, J. B., & Csibra, G. (2018). Why do we remember? The communicative function of episodic memory. *Behavioral & Brain Sciences*, *41*. doi: 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. doi: 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*(3), 367–380. doi: https://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. doi: 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*, 409–414. doi: https://doi.org/10.1177/0963721419847998

Maynard Smith, J., & Harper, D. (2003). *Animal signals*. Oxford University Press.

McDaniel, M. A., & Einstein, G. O. (1986). Bizarre imagery as an effective memory aid: The importance of distinctiveness. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *12*(1), 54–65. doi: https://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. doi: http://dx.doi.org/10.1037/gpr0000111

Mercier, H. (2020). *Not born yesterday: The science of who we trust and what we believe*. Princeton University Press.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral & Brain Sciences*, *34*(2). doi: https://doi.org/10.1017/s0140525x10000968

Mercier, H., Majima, Y., & Miton, H. (2018). Willingness to transmit and the spread of pseudoscientific beliefs. *Applied Cognitive Psychology*, *32*(4), 499–505. doi: https://doi.org/10.1002/acp.3413

Mercier, H., Majima, Y., Claidière, N., & Léone, J. (2019). Obstacles to the spread of unintuitive beliefs. *Evolutionary Human Sciences*, *1*(E10). doi: https://doi.org/10.1017/ehs.2019.10

Mermelstein, S., Barlev, M., & German, T. C. (2019, February). *Tell me more about that melting lizard! Counterintuitive concepts trigger information search*. Paper presented at the Evolutionary Psychology Pre-Conference at the annual Convention of the Society for Personality and Social Psychology, Portland, OR.

Mermelstein, S., Barlev, M., & German, T. C. (2020). She told me about a singing cactus: Counterintuitive concepts are more accurately attributed to their speakers than ordinary concepts. *Journal of Experimental Psychology: General*, *150*(5), 972–982. doi: https://doi.org/10.1037/xge0000987

Mermelstein, S. & German, T. C. (2021). Counterintuitive pseudoscientific beliefs propagate by exploiting the mind's communication evaluation mechanisms. *Frontiers in Psychology*, *12*, doi: https://doi.org/10.3389/fpsyg.2021.739070

Mermelstein, S., & Barlev, M., Alrifai, A., & German, T. C. (in prep). *Cultural inputs modulate but do not revise core intuitions: Representational co-existence in Christian and Islamic God concepts*. Preprint: https://osf.io/g3uv5/

Mills, C. M. (2013). Knowing when to doubt: developing a critical stance when learning from others. *Developmental Psychology*, *49*(3), 404–418.

Miton, H., & Mercier, H. (2015). Cognitive obstacles to pro-vaccination beliefs. *Trends in Cognitive Sciences*, *19*(11), 633–636. doi: https://doi.org/10.1016/j.tics.2015.08.007

Miton, H., Claidière, N., & Mercier, H. (2015). Universal cognitive mechanisms explain the cultural success of bloodletting. *Evolution & Human Behavior*, *36*, 303–312. doi: http://dx.doi.org/10.1016/j.evolhumbehav.2015.01.003

Moore, D. W. (2005, June 16). Three in four Americans believe in paranormal. *Gallup*. Retrieved from https://news.gallup.com/poll/16915/three-four-americans-believe-paranormal.aspx

Morin, O. (2013). How portraits turned their eyes upon us: Visual preferences and demographic change in cultural evolution. *Evolution & Human Behavior*, *34*, 222–229. doi: http://dx.doi.org/10.1016/j.evolhumbehav.2013.01.004

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

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

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

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, *25*(5), 388–402. doi: https://doi.org/10.1016/j.tics.2021.02.007

Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences*, *107*(Supplement 2), 8993–8999. doi: https://doi.org/10.1073/pnas.0914630107

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. doi: http://dx.doi.org/10.1016/j.vaccine.2010.02.052

Reber, A. S., & Alcock, J. E. (2020). Searching for the impossible: Parapsychology's elusive quest. *American Psychologist*, *75*(3), 391–399. doi: https://doi.org/10.1037/amp0000486

Richter, T., Schroeder, S., & Wöhrmann, B. (2009). You don't have to believe everything you read: Background knowledge permits fast and efficient validation of information. *Journal of Personality & Social Psychology*, *96*(3), 538–558. doi: https://doi.org/10.1037/a0014038

Shtulman, A. (2013). Epistemic similarities between students' scientific and supernatural beliefs. *Journal of Educational Psychology*, *105*(1), 199–212. doi: https://doi.org/10.1037/a0030282

Shtulman, A. (2017). *Scienceblind: Why our intuitive theories about the world are so often wrong*. New York: Basic Books.

Shtulman, A., & Harrington, K. (2016). Tensions between science and intuition across the lifespan. *Topics in Cognitive Science, 8*, 118–137. doi: https://doi.org/10.1111/tops.12174

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, *124*, 209–215. doi: https://doi.org/10.1016/j.cognition.2012.04.005

Singh, M. (2018). The cultural evolution of shamanism. *Behavioral & Brain Sciences*, 41(E66). doi: https://doi.org/10.1017/S0140525X17001893

Slone, J. (2007). *Theological incorrectness: Why religious people believe what they shouldn't.* Oxford University Press.

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

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

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, *99*(4), 605–632. doi: https://doi.org/10.1037/0033-295X.99.4.605

Sperber, D. (1985). *On anthropological knowledge: Three essays*. New York: Cambridge University Press.

Sperber, D. (1994a). Understanding verbal understanding. In J. Khalfa (Ed.), *What is intelligence?* (pp. 179–198). Cambridge University Press.

Sperber, D. (1994b). The modularity of thought and the epidemiology of representations. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 39-67). Cambridge: Cambridge University Press.

Sperber, D. (1996). *Explaining culture: A naturalistic approach*. Blackwell.

Sperber, D. (1997). Intuitive and reflective beliefs. *Mind and Language*, *12*, 67–83. doi: https://doi.org/10.1111/j.1468-0017.1997.tb00062.x

Sperber, D (Ed.). (2000). *Metarepresentations: A multidisciplinary perspective*. Oxford University Press.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition*. Wiley-Blackwell.

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

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*(6230), 91–94. doi: https://doi.org/10.1126/science.aaa3799

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*(1), 42–61. doi: https://doi.org/10.1037/0033-2909.111.1.42

Tooby, J. & Cosmides, L. (1992). The psychological foundations of culture. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 19–136). Oxford University Press.

Tooby, J., & DeVore, I. (1987). The reconstruction of hominid behavioral evolution through strategic modeling. In W.G. Kinzey (Ed.), *The evolution of human behavior: primate models* (pp. 183–237). State University of New York Press.

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

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

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, *4*(1), 45–69. doi: https://doi.org/10.1016/0959-4752(94)90018-3

Wertz, A. E. (2019). How plants shape the mind. *Trends in Cognitive Sciences*, 23, 528–531. doi: 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. doi: http://dx.doi.org/10.1177/0956797613516145

Whitson, J. A., & Galinsky, A. D. (2008). Lacking control increases illusory pattern perception. *Science*, *322*(5898), 115–117. doi: https://doi.org/10.1126/science.1159845

Wood, M. J., Douglas, K. M., & Sutton, R. M. (2012). Dead and alive: Beliefs in contradictory conspiracy theories. *Social Psychological & Personality Science*, *3*(6), 767–773. doi: https://doi.org/10.1177/1948550611434786

Zahavi, A. & Zahavi, A. (1997). *The handicap principle*. New York: Oxford University Press.

APPENDICIES

Appendix 1: Full set of concepts

Full set of counterintuitive and ordinary concepts

Noun Pairs	Domain	Counterintuitive descriptor	Breach / Transfer	Violation
Cat / Dog	Animal	that has brown spots and can walk through solid walls	Breach	Physics
Beetle / Earthworm	Animal	that hides in a log and knows everybody's inner-most thoughts	Breach	Psychology
Lizard / Rat	Animal	that has a long, thin tail and can never die no matter what happened to it	Breach	Biology
Table / Chair	Artifact	that is big and often floats in midair	Breach	Physics
Fence / Mailbox	Artifact	that is covered with moss and is crying because it is sad	Transfer	Psychology
Hammer / Shovel	Artifact	that has a wooden handle and needs food every day to stay strong	Transfer	Biology
Branch / Rock	Object	that feels cold to the touch and can speak in French	Transfer	Psychology
Cloud / Rainbow	Object	that is large and far away, and knows what happens in the future	Transfer	Psychology
River / Mountain	Object	that is near a forest and can overhear everything people say by it	Transfer	Psychology
Shrub / Cactus	Plant	that is small in size and likes to sing loudly	Transfer	Biology
Banana / Mango	Plant	that is very fresh and ripe and turns invisible a few minutes every day	Breach	Physics
Rose / Tulip	Plant	that sways in the wind and can be in two different parts of the world at the exact same time	Breach	Physics

Noun Pairs	Domain	Ordinary descriptor
Cat / Dog	Animal	that has soft fur and likes to play with toys
Beetle / Earthworm	Animal	that stays in the mud and moves slowly inch by inch
shrub / Cactus	Animal	that eats insects off the ground and moves around quickly on all four of its feet
Table / Chair	Artifact	that is firm to the touch and can hold lots of weight
Fence / Mailbox	Artifact	that is made of metal and has sharp edges along its corners
Hammer / Shovel	Artifact	that feels heavy to hold and is a light brown in color
Branch / Rock	Object	that is thick and hard and looks shiny in the sunlight
Cloud / Rainbow	Object	that is high in the sky and can be seen from the ground
River / Mountain	Object	that takes time to cross over and is surrounded by cold, clear air
Shrub / Cactus	Plant	that is dark green and is growing next to a stream
Banana / Mango	Plant	that has a bright yellow skin and a very fruity and delicious smell
Rose / Tulip	Plant	that has roots that go deep into the soil and need sunshine to grow or else it would be small

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

Appendix 2: Story Stimuli (Person condition, stimuli list 1 from Exp. 1)

- *Story 1*

[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 on 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.

- *Story 2*

[Miguel / Joanna / Sam / Ariel] tells you the following story:

A man went on many road trips all around the country. He had to travel so much because he worked for a large company. He had to go around to different stores and fix things when they broke. One day, after he had come back after a long journey, he told his friends about what he had seen along the way.

While he was driving, he saw a rock on a nearby road. This was a rock that was thick and hard and looked shiny in the sunlight. After driving a little further, he found himself passing through a small woody area. When he was deep in the woods, he saw a tulip planted in the ground. The tulip had roots that went deep into the soil and needed sunshine to grow or else it would be small. At the end of the woods, the man took a break from driving and came across a banana lying on top of a pile of leaves. The banana was very fresh and ripe and turned invisible a few minutes every day.

The man was done driving through the woods, but next had to drive through a couple towns. At one point he drove by a fence at the intersection of two roads. This was a fence that was covered with moss and was crying because it was sad. He was getting hungry now, because he had been driving for a long time. He decided to drive a little faster to make it home quicker. When he reached his neighborhood he saw a dog in the front yard of a neighbor's house. The dog had soft fur and liked to play with toys. Finally, the salesman arrived at the front door of his house. As he pulled into his driveway, he saw a branch just above his head on the roof of his garage. The branch felt cold to the touch and could speak in French.

The salesman was happy to be home. He couldn't wait to tell his friends about what he had seen on his travels.

- *Story 3*

[Miguel / Joanna / Sam / Ariel] tells you the following story:

Dr. Wurg was recently selected to be the new ambassador to a distant country. To prepare for this new assignment, Dr. Wurg decided to visit a local museum which had exhibits on representative objects and animals from this country. After a short train ride, Dr. Wurg arrived at the museum. She had a coffee at the museum café, and then was ready to explore the different halls of the museum to see what life is like in the country she's going to.

The first room she went into had information about different types of plants. A cactus in the corner of the room caught Dr. Wurg's attention. This cactus was small in size and liked to sing loudly. After enjoying that exhibit, Dr. Wurg looked out the window and saw a river. She knew that this river took time to cross over and was surrounded by cold, clear air. Dr. Wurg then decided to have a mango as a snack. This was a mango that had a bright yellow skin and a very fruity and delicious smell. The day was growing short so Dr. Wurg decided to try to see more of the museum before it closed.

Walking down the hallway away from the window, Dr. Wurg came across a cat. The cat had brown spots and could walk through solid walls. Next there was an exhibition about the history of the country where Dr. Wurg was going. Dr. Wurg was very interested in a table that was on display. This table was big and often would float in midair. Since her time was running short, Dr. Wurg quickly visited the rest of the museum. On her way out of the

museum, Dr. Wurg noticed a beetle. The beetle stayed in the mud and moved slowly inch by inch.

Reflecting on the museum on the train back home, Dr. Wurg felt that she learned a lot about the country she would be living and working in.

- *Story 4*

[Miguel / Joanna / Sam / Ariel] tells you the following story:

A father and son decided to go on a hike up on some hills. They woke up very early in the morning and packed everything they would need. They packed water, food, and extra socks in case they wore out the socks they were wearing. When they were ready to go, the father started up the car, and they drove up into the hills.

As they were driving the boy looked out the window of the car and saw a mailbox stuck in the ground. The mailbox was made of metal and had sharp edges along its corners. The father and son then arrived at the start of the hiking trail. They parked the car and began to walk. The father pointed out a cloud to his son which was high up in the sky. The cloud was large and far away, and knew what happened in the future. The father and son continued their hike, walking up and down many different hills. Hiking was tough work and they were both getting tired, so they stopped to rest by a shrub. The shrub was dark green and was growing next to a stream.

After a short rest the father and son continued on with their hike. From the top of one hill they could see a nearby mountain. The mountain was near a forest and could overhear everything people said by it. The sun was beginning to set so they turned around and began the hike back to the car. On the way back they noticed an earthworm. The earthworm hid in a log and knew everybody's inner-most thoughts. Finally, they made it back to their car and drove home. Upon returning home and walking into the kitchen, the son saw a chair which had not been there before. The chair was firm to the touch and could hold lots of weight.

It rained that night. The father and son were tired from their long day up in the hills, but they were very happy with all that they've seen.