ABSTRACT

Title of Dissertation:	REFRAMING CHILDREN'S JUDGMENTS OF CONSENSUS RELIABILITY AS A PROCESS OF INFORMATION AGGREGATION
	Karen Carmel Levush, Doctor of Philosophy, 2023
Dissertation directed by:	Associate Professor, Lucas Payne Butler, Department of Human Development & Quantitative Methodology

Consensus is a compelling cue to the truth value of a given claim, but certain consensus patterns provide stronger evidence than others. This dissertation examines the developmental trajectory of children's reasoning about the epistemic value of diverse perspectives for consensus' reliability. One-hundred forty-four children between the ages of 7 and 9, as well as 48 adults, were introduced to a novel planet and alien groups that live there. Tasked with learning the "right things" about why various natural phenomena occur on this planet, participants were asked which one of two consensus groups, each of whom collectively thought something different, was the "better" group to ask. Participants rated their relative preference for one consensus group over another using a 6-point scale and were asked to explain their reasoning. These findings provide initial evidence that qualitative changes in children's ability to consider how dependencies can lead to redundant information parallel the developmental shift in children's appreciation for interpretive diversity in middle childhood.

REFRAMING CHILDREN'S JUDGMENTS OF CONSENSUS RELIABILITY AS A PROCESS OF INFORMATION AGGREGATION

by

Karen Carmel Levush

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:
Associate Professor, Lucas Payne Butler, Chair
Professor, Geetha Ramani
Professor, Kathleen Corriveau
Associate Professor, Richard Prather
Associate Professor, Yi Ting Huang

Dedication

To my Ima and Aba,

for your patience, wisdom, and constant flow of love and support.

Acknowledgements

I am immensely thankful for all the individuals who have supported me throughout this journey, including, but surely not limited to:

The chair of my dissertation committee and PhD advisor, Lucas Payne Butler, for taking a chance on a recovering special education teacher, welcoming me into your research lab, providing me with thoughtful feedback throughout my time in the program, and continuously pushing me beyond what I thought I was capable of.

My dissertation committee members, Kathleen Corriveau, Yi Ting Huang, Richard Prather, and Geetha Ramani, for your invaluable feedback and time dedicated to serving on my committee.

Mary DePascale, for your selfless support, feedback, and weekly working sessions that have kept my dissertation writing on track.

Hailey Gibbs, for your friendship (that began when I started the master's program in 2016) that supported me throughout each stage of this long academic journey.

Current and former members of the Cognition and Development Lab, Emily Berman, Spencer Lin, Karmen Macchiagodena, Nicole Calianese, Jackson Eisner, Emily Mitchener, Teresa Garcia, and Jenna Alton, for your discussion and help with data collection, transcription, coding, and much more.

HDQM Faculty, for creating a welcoming and rigorous learning environment.

My sisters, for your ongoing love and support, long phone calls and (hopefully) excusing my absence at birthday parties and other get-togethers when I was stressed with deadlines.

Julia Merlin, for serendipitously moving in across the street, becoming friends, and helping me with last minute statistics questions.

Online PhD meme-makers, for making me laugh between marathons of reading and writing.

My husband, Alex, for being my rock and lifelong partner, in-person screen reader, readability editor, and letting me rant about things completely outside of your area interest.

"*Children are socialized and integrated into their school environment that often mirror the social structures within their larger cultural context…My interest in the dynamic interplay between children's developing understanding of social structures, motivation to learn, and their cognitive development is inspired by my experiences in the classroom*."

- Me, 2017 (UMD PhD Personal Statement)

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Statement of the Problem

We are all dependent on testimony for much of what we know about the world (Coady, 1992). But we also frequently receive testimony from multiple sources *on the same issue*. One cue we commonly use to modulate our confidence in such cases is the degree of consensus between sources. Consensus is a compelling cue to the truth-value of a given claim; learners — across cultures (Boehm, 1996; van Leeuwen et al., 2018), age groups (Haun & Tomasello, 2011; Mannes, 2009), and species (Haun et al., 2013; Claidière & Whiten, 2012) — are generally inclined to adopt a view endorsed by a majority. But there are also many situations we may face in which there is no clear majority, rather there are multiple and even contradictory viewpoints, each of which may be shared by different groups of people. The individuals comprising real-world consensus groups may be independent of one another, producing data and conclusions in isolation from other consensus members with whom they agree. However, in many if not most circumstances, they will display some degree of dependence: they may share a common background, share a common source of evidence, or even discuss their evidence with others prior to offering their individual reports.

A central feature of testimony that helps us to determine its truth is the reliability of its source. Thus, when accepting others' testimony, we endorse — at least tacitly — the process by which the informant(s) came to believe the claim asserted. This *perceived reliability* determines how much weight the source's testimony is assigned. Decades of developmental research on trust and testimony have shown the many ways in which even preschoolers are skilled at evaluating the reliability of sources using both *direct cues* to sources' knowledgeability, like prior accuracy

or lack of information access (e.g., Koenig & Harris, 2005; Robinson, Champion & Mitchell, 1999), as well as *indirect cues* such as their broader social networks and their position within it (e.g., Kinzler et al., 2011; Landrum et al., 2013; Sóley, 2019). Study findings related to children's use of *direct cues,* in particular, provide key insights into how children leverage their developing understandings of how certain experiences lead to certain beliefs to evaluate the reliability of an informant and the testimony they provide. The importance of this work here is that it illustrates how children begin to appropriately integrate information about *single informants* over the course of *multiple* events to make these inferences. Crucially, and of particular importance to the current dissertation, this framework can also be applied to studying how they may integrate information about *multiple informants* for *one* event: a consensus.

Past research investigating the developmental origins of the tendency to look for and follow a consensus by and large has focused on contexts in which there is a clear majority who are in agreement (e.g., Haun et al., 2012; Fusaro & Harris, 2008; Morgan et al., 2015). While multiple converging reports can provide stronger evidence than one informant alone, there are many possible routes by which people can come to share beliefs about the world. Critically, the strength of the support offered by converging sources is dependent on how their reports are aggregated. For example, if we find that multiple converging reports rely on the same *shared* evidence, then the strength of the support offered by the consensus *itself* is dependent on that evidence. Such a dependency results in redundant information across reports, reducing evidential support, as compared to multiple converging reports based on distinct pieces of evidence. In everyday life, however, it is a common feature of real-world testimony that a recipient does not know exactly *how* individual sources come to believe what they believe and report to others—

rather, the only source of evidence available may be specific patterns of variation and covariation between the beliefs people hold and with whom they affiliate.

Study Rationale

The current study addresses a number of gaps in the developmental literature. By leveraging a diverse set of findings including testimony and argumentation, inductive reasoning, and social categorization, the following connections provide a foundation that motivate this dissertation.

Social Affiliation as Cue to Dependence

The human social world is comprised of a complex web of social relations; people belong to many different social groups at the same time. The interrelation between any given society's networks can represent overlapping channels of information. This means that how we represent not only our relationships with others, but also others relationships and the social networks that those relationships comprise, can be used to infer dependencies. Thus, social connection on its own can undermine the perceived reliability of a claim when multiple converging reports may originate from a shared source. It can also lead to biased decision making when representations of sources' social networks are inaccurate or based on stereotypes.

Tracking others' relationships is therefore a critical skill, as it enables children to make inferences about who is likely to help them acquire what knowledge (e.g., Fitneva et al., 2013; Sóley, 2019). The ability to effectively navigate our complex social environment relies on our capacity to form a representation that accurately reflects the real world. Research in adult social psychology shows that humans build flexible "mental maps" to mentally array groups on at least two or more dimensions (Fiske et al., 2002). Recent pioneering work in social neuroscience

suggests that similar to how we navigate physical space, the way we make social decisions also reflects a kind of navigation through abstract social dimensions (see Park et al., 2021). There are infinite dimensions in which people can be grouped, all of which comprise a basis for which social relations can be encoded and used to infer why certain patterns of behavior are observed. Further, these dimensions vary across a continuum of abstraction. At a more concrete level, the relationship between two people that sit next to each other at a bus stop in the morning can be defined by the physical proximity between their homes. At the same time, at a more abstract level, their relationship might be further defined by the fact that they are both elementary school teachers. Both the proximity and social group categorizations can be used to make deductive inferences about why they are sitting at a bus stop in the morning.

The finding that children's abstract representations of social groups include expectations that categories constrain behavior (and thus experiences) is well-documented, for both familiar social categories such as gender and novel groups to which they themselves do not belong (e.g., Blakemore, 2003; Gelman et al., 2010; Liben et al., 2002; Rhodes & Chalik, 2013; Roberts et al., 2017; Shutts et al., 2010, 2013). Given the interrelatedness of the social world, the extent to which channels of information overlap can be used as a justification for supporting a view endorsed by a consensus. Whether relations between consensus members exist is not necessarily the most important question, rather it is whether their overlapping networks reflect privileged access to knowledge and, secondly, the number of unique, diverse routes for knowledge acquisition.

While there is a large body of work examining how children come to understand the value of diversity for drawing broader generalizations (e.g., Gutheil & Gelman, 1997; Li et al., 2009; Rhodes & Brickman, 2010; Rhodes et al., 2008), no research has directly examined

children's representations of social diversity and how they may reason about its epistemic value. Given this, the importance of exploring whether and how children might use different patterns of consensus to infer that a certain position is correct is an essential next step.

Direct vs. Indirect Cues to Dependence

Developmental trends in reasoning about how dependencies between consensus members and their testimonial reports, particularly in situations where there are an equal number of endorsements for different positions, has received surprisingly little attention. While there are few studies that are relevant to this issue, this work has focused exclusively on children's use of *direct* cues to (in)dependence between sources; that is, *how* consensus members came to believe what they agree upon is discernable through observable behaviors (e.g., looking inside a box) or made explicit (e.g., "Peter told me"; "I saw"). Results from this work suggest that only by middle childhood do children demonstrate a stable understanding that multiple reports originating from the same or shared information compromises its evidential value (e.g., Aboody et al., 2022). Given that we often do not have access to *direct* knowledge of how consensus members come to agreement, this dissertation seeks to build a bridge from prior work on how children aggregate information from multiple sources using direct cues to dependence, to children's use of more commonly available *indirect* cues: social relations.

Group Membership as a Cue to Shared Knowledge

More research is needed to explore the impact of children's developing ability to reason about the link between peoples' experiences and the beliefs that they hold, across different levels of abstraction. The central thesis of this dissertation is that children's understanding that different experiences lead to different knowledge-states and their representations of social categories, can

shape how they reason about different patterns of consensus as a source of evidence for the truth or falsity of underlying claims. In the context of testimonial learning, children's use of social category information – such as gender (Shutts et al., 2010; Taylor, 2013), race (Chen et al., 2013), language, (Kinzler et al., 2007), accent (Kinzler et al., 2013), and even minimal group membership (Buttelmann et al., 2013) – to guide their trust are commonly framed as evidence of an early emerging tendency to privilege information from ingroup members, as opposed to outgroup members. Nonetheless, a growing body of work suggests an alternative account, namely, that social category information can serve an epistemic function linking affiliative relations between individuals to an expectation of *shared knowledge* (see Oláh et al., 2019).

Social categorization in children has been extensively investigated both in relation to the social categorization process itself and the (affective) implications for others' interactions and behaviors (see e.g., Liberman et al., 2017; Liberman & Shaw, 2020). For the purpose of the current dissertation, however, the focus is on the intersection between representations of social categories and conceptual changes in children's intuitive "Theory of Mind" through middle childhood -- from first recognizing there is variation in people's beliefs about the world, to then appreciating differences in people's experiences as the source of that variation, and finally to reasoning about how different types of knowledge are distributed in others' minds.

However, the human social world is comprised of a complex web of social relations and hierarchies. Thus, children must use their knowledge of others' social relations flexibly in determining the significance of different social category markers in social learning. Nevertheless, and as emphasized in every statistics course, correlation is not *causation*; if misapplied, such assumptions about the causal power of mere membership of social categories for knowledgeability can lead to dangerous over-weighting of evidential value. So, children must be

able to keep track of multiple hypotheses and their associated probabilities to selectively learn from reliable sources. This makes it essential to investigate how children integrate information about multiple sources, and how they accommodate a key feature of multiple sources in the real world, namely that these sources may not be entirely independent.

Current Study

The goal of this dissertation is to examine developmental trends in reasoning about *indirect* dependencies between consensus members introduced through sharing common backgrounds, and its impact on evaluations of consensus' accuracy. Focusing on contexts for which there is no available information as to how each group's opinion is formed, this dissertation investigates how children integrate information about multiple converging sources based on cues to indirect dependence at different levels of abstraction. Specifically, the dimensions of social group membership (*shared* vs. *distinct*) and proximity (*close* vs. *distant*) are manipulated with the goal of examining how children represent, reason, and integrate each of these cues as constraints on the experiences underlying agreement between consensus members. The hope is that this work contributes to a broader conception of how an understanding of the epistemic value of social diversity itself develops and is leveraged for evaluating consensus accuracy.

Aims & Hypotheses

Aim 1. Examine children's sensitivity to diversity based solely on social group membership. The preliminary studies that form the basis of this dissertation suggest that it is not until 8 years of age that children can recognize the merits of diverse perspectives for forming an objective viewpoint (Levush et al., in prep, see Ch. 2 for further discussion). A key limitation in

that work is that proximity and social group membership were confounded. Thus, for the purposes of the present study, the first aim was to investigate whether disentangling these two indirect cues to dependence would result in consistent findings. If both cues are represented in terms of shared knowledge or experience, then we can expect the finding that children around age eight begin to prefer diversity among consensus members holds even when there is no correspondence between group membership and proximity.

Aim 2. Examine developmental change in children's preference for social diversity between converging sources. This aim addresses the need to track whether the developmental shift in children's preferences for social diversity among consensus members reflects an emerging ability to consider the epistemic value of diversity *itself* between converging sources.

If children represent social group membership in terms of overlapping experiences, developmental change in children's appreciation of how social diversity generates epistemic value should correlate with an increasing sensitivity to both the *distant* proximity and the *distinct* group membership consensus group properties. Previous research indicates that the different social dimensions along which we form judgements about others is hierarchically organized, that is, we prioritize certain dimensions of social categorization over others (e.g., Kinzler et al., 2010). Given that it comprises of the relation between consensus members' physical locations, *close*/*distant* proximity relations between consensus members imply opportunities for overlap in experiences in a more concrete way than that of group membership. When members of a consensus with *distinct* group membership are of *close* proximity, it was expected that children's relative preference for the consensus with *distinct* group membership would shift in the opposite direction.

Aim 3. Examine the explanatory process underlying children's consensus group preferences. There can be different reasoning underlying the same patterns of responses across relative consensus group preferences. Specifically, the preferences listed in Aims 1 and 2, can be based on more associative grounds, or alternatively, a more *causal* understanding of social diversity as a mechanism for consensus accuracy. An associative account could simply be based on the fact that greater social diversity is more representative of the target population for which the knowledge in question is expected to be distributed. Whereas a more sophisticated understanding can reflect a sense of how social diversity generates epistemic value.

There are several distinct ways to think of the benefits of diversity for consensus accuracy, all of which correspond to abstract relations between diversity as a property of a consensus group and the experiences that underlie agreement among them. These accounts fall broadly into three types. The first is to think of diversity as a proxy for unique sources of evidence, where the diversity of consensus members corresponds to the distinct sources of experiences. When characterized in this way, a diverse consensus can be thought of as comprising distinct pieces of knowledge that, when combined, can be used to generate a reliable approximation of the truth, or alternatively help rule out a greater number of potential explanations. The second way requires considering the normative dimension of knowledge. By thinking of diversity in terms of a reflection of knowledge propagation, the less overlap between social networks the more widespread a belief is, thus reflective the normative truth of a claim. The third way appeals to the social processes underlying consensus formation; this view involves probabilistic inferences about potential shared sources of consensus members' judgment errors – namely, those that are systematic or biased towards a particular direction. Critically, this third

way is analogous to the notion of consilience, or that different methods studying the same phenomena should produce converging evidence.

Chapter 2: Review of the Literature

Overview

This chapter provides a theoretical and empirical background on how children use their developing ability to integrate information from multiple sources, in early and middle childhood, in order to make sense of people's divergent perspectives. Specifically, I outline a framework for charting qualitative developmental changes in children's judgments of reliability centered on dependency structures. As described in Chapter 1, the overarching hypothesis of this current dissertation research is that these dependency structures underlie many (perhaps even all) judgments of reliability. In this chapter, I focus on showing how this perspective can be used to build a bridge between the findings on children's reasoning about the reliability of single informants and how they evaluate consensus accuracy.

First, relevant literature on judgments of reliability, including direct and indirect ways to judge expertise and trustworthiness are discussed. Second, I provide background on understanding experience as the generative source of a belief, including a discussion of Theory of Mind, definitions of sources and domains of knowledge, and information about domain-dependent distributions of knowledge. Third, I discuss how children's judgments of reliability can be reframed as a process of information aggregation. Finally, I discuss children's evaluation of evidence based on probabilistic relations and the influence of dependencies on judgments of reliability, including evidence from my own recently-completed preliminary studies, which shape the methods and aims of the current dissertation.

Research examining developmental change in children's ability to reason about the reliability of sources has generally focused on early childhood and been studied through single

informant paradigms. Whether and when children begin to demonstrate sensitivity to dependencies between consensus members, particularly in situations where there are an equal number of endorsements for different positions, has received surprisingly little attention. However, there are a couple of recent studies that are relevant to this issue, and as such, are presented in this review. Importantly, each study using a single source paradigm is coupled with an analogous study in which children evaluate claims from multiple converging sources. Given that knowledge attributions are domain dependent, the analysis of the relation between single and multiple informant paradigms should be reserved to when the claims in question are of the same domain. Further, the aim of presenting these studies coupled in this way is to highlight the dependency as the common structure from which children integrate information about informants' reliability, which is the focus of the current dissertation.

Judgments of Reliability: Foundational Concepts

Central features of testimony that help us to determine its truth are the expertise and trustworthiness of the source, as well as the extent to which its contents fit with our own understanding (Hahn et al., 2016). Evaluations of an informant's "expertise" are an assessment of their knowledge-state, and thus signify their *capacity* to provide accurate information about a given topic. In other words, evaluations of an informant's expertise involve attributions of the *kinds* of experiences that lead to knowledge and thus are domain dependent. For example, a music teacher can provide relevant and accurate information about types of instruments but may not be able to provide guidance on construction of office buildings. Trustworthiness, on the other hand, refers to the *intention* to provide true and accurate information to the best of one's ability. For example, if a classmate has a motive to attract other classmates to play on a particular playground after school, they may falsely claim that the playground has more swing sets than

any other playground in the neighborhood, even if it in fact has just as many. But there are also cases that are more subtle; someone can have every intention to provide objective judgements but can be challenged in doing so for a variety of reasons, including having limited access to the range of information available that is relevant to make an informed judgement or having implicit biases based on preexisting beliefs or preferences that lead to skewed interpretations of data.

In relation to evaluations of expertise, there is a great deal more work dedicated to how children integrate information that *directly* bears on informants' expertise, such as their track record of providing accurate testimony (Birch et al., 2008; Jaswal & Neely, 2006; Koenig et al., 2004; Koenig & Harris, 2005; Koenig & Woodward, 2010; Scofield & Behrend, 2008; Pasquini et al., 2007), whether they have access to a relevant epistemic source (Koenig, 2012; Robinson et al., 2011; Brosseau-Liard & Birch, 2011; Nurmsoo & Robinson, 2009), and whether they are labeled an expert of a relevant domain (Kushnir et al., 2013; Lane & Harris, 2015; Lutz & Keil, 2002). Of course, there are also many *indirect* ways by which informants' expertise can be inferred, such as on the basis of specific features or properties of their immediate environment (Lockhart et al., 2016), their interpersonal relationships, and broader social networks (e.g., Liberman et al., 2019). For example, someone who lives in Hawaii may be more likely to know that the island chain was formed by volcanoes than someone who lives in Bulgaria.

As for discerning the trustworthiness of a source, this can be more challenging. It is certainly possible for someone to have knowledge and not want to use it, and it is also possible for someone to have access to an epistemic source and choose not to leverage that access for a wide range of reasons. Critically, the choice of not leveraging access to an epistemic source can be warranted. For example, if a student seeks help from their friend who they know has strong understanding of fractions, but that friend has a more pressing assignment due in another subject,

they may regrettably let the student know they are unable to help at that time. A choice of not leveraging access to an epistemic source also need not be a calculated intentional act, but rather an implicit decision that is made. Of central importance here, while an informant may have access to an epistemic source in principle, recognizing how their social relationships might limit their awareness of that access is key to developing an understanding of how judgements can be biased in more subtle ways than outright mal intent.

Expertise and trustworthiness are distinct concepts such that an informant can be highly expert but highly untrustworthy, and vice versa. Nevertheless, these two factors ultimately combine to determine how likely it is that a source's testimony genuinely reflects the truth or falsity of a claim. The degree to which we believe that a source has the capacity and willingness to provide accurate information about a given claim at issue is defined as their overall 'reliability'. This *perceived* reliability determines how much weight the source's testimony is assigned, or in other words, it determines the evidential value assigned to that report. Nevertheless, an essential foundation for recognizing what information is relevant for evaluating informants' expertise and trustworthiness is an ability to appreciate that what people know and believe are shaped, and constrained, by their experiences.

Understanding Experience as the Generative Source of a Belief

Estimating another person's knowledge-state is a key component of human Theory of Mind and critical for building on the expertise of others. An understanding that experience is a *generative* source of a belief facilitates an ability to not only appreciate that peoples' beliefs can vary, but to reason about differences in peoples' experiences as the source of that variation (e.g., Gopnik & Astington, 1988; Eisbach, 2004). An understanding that peoples' beliefs about the world can vary is foundational for recognizing that different pockets of expertise can be

associated with different individuals. However, an ability to consider the kinds of experiences that lead to expertise relies on representations of what constitutes knowledge in the first place.

The classic false-belief task, first introduced as a measure of children's Theory of Mind development, was based on this very premise, namely that children will only be able to attribute a false belief to an actor once they recognize the deceived actor's perceptual experience as the source of their false belief. The classic measure of a child's false-belief understanding is the change-of-location task (Wimmer & Perner, 1983) in which a child is read a story about a character named Sally, who stores her toy and then leaves the room. While she is away, another character moves the toy to a different location. Sally then returns to search for her toy and the child is asked, "Where will Sally first go look for her toy?" Decades of research have suggested that 3-year-olds tend to struggle in their reasoning in a specific way— they say that Sally will point to the location where the toy is currently sitting, which is the location consistent with reality, rather than the location consistent with Sally's false-belief. In contrast, 4-year-old's more often appropriately respond that Sally will look in the location where she initially left the toy, indicating that they can successfully considering an agent's beliefs (e.g., Baron-Cohen et al., 1985; Perner, et al., 1987; Wimmer & Perner, 1983).

The developmental achievement of false belief understanding is linked to children's ability to recognize perceptual experience as a source of knowledge, and thus an understanding of beliefs as representations of reality rather than realities themselves (e.g., Gopnik, 1993; Perner et al., 1987; Wellman, 1990; Wellman et al., 2001). However, not all knowledge can be learned through direct observation (i.e., perceptual experience); people can also learn from communication with others and from logical inference given the available cues in a situation. Critically, the capacity for both these forms of knowledge acquisition is likely to be more

variable across individuals than is the capacity to learn from perception. Children's ability to recognize the range of learning methods possible is an essential foundation for discerning which informants are more likely to provide accurate information. But it crucially relies on more advanced understanding of the constructivist nature of beliefs—specifically, an appreciation for *interpretive diversity*, that there can be different interpretations of the same reality, or in other words, people can have varying perspectives.

Consider the scenario where two individuals both look a painting of a ship in rough seas. One of the individuals may interpret the piece of art as highlighting the significance of persistence, while the other individual may perceive it as emphasizing the role of destiny. Both interpretations are valid and neither is objectively true or false. Without an understanding of interpretive diversity, a child would not recognize that multiple perspectives on this piece of art can coexist. Rather, they would assume that both individuals would share the same belief about its meaning simply because they looked at the same painting.

It is precisely *because* variation in people's experiences may lead to differences in what they believe that it is therefore possible for people to have different interpretations of the same event (i.e., interpretive ToM; see Carpendale & Chandler, 1996). Unlike false-belief understanding, there is no gold standard method of assessing more advanced ToM, rather researchers have used a variety of tasks to investigate later developments. Broadly, this research suggests it is not until middle childhood, around 7 or 8 years, that children come to understand that differences in beliefs arise from differences in *perspective* (e.g., Chandler & Carpendale, 1998; Lalonde & Chandler, 2002; Ross et al., 2005). An understanding that different beliefs can arise solely from differences in interpretative mental processes enables children to appreciate that

knowledge can be more or less certain, as well as recognize some sources of knowledge as better than others (Heiphetz et al., 2013; Kuhn, 2020; Kuhn et al., 2000).

Conception of Knowledge. For the purposes of this dissertation, knowledge refers to a justified true belief. The term "justified" is referenced here as a property of a belief, and thus necessitates both normative and evaluative concepts of epistemic standards (Chignell, 2010). Epistemic standards refer to the criteria used to evaluate the credibility of a source of information, such as what kind of evidence is good evidence in a particular context (Chinn et al., 2011). Judgements of reliability, therefore, refers to an evaluative process through which information deemed relevant for the question at hand is evaluated against epistemic standards.

Many studies have shown children's use of epistemic standards, though this point is not generally highlighted in the extensive research on children's understanding of belief (Wellman, 2014). Nevertheless, the epistemic standards that individuals employ change developmentally (see Iordanou, 2016). Insights from research on Theory of Mind, as well as that of testimony and argumentation, contribute to addressing the question of what develops in children's ability to aggregate information from multiple sources in order to make sense of people's divergent perspectives and make judgments of reliability.

Knowledge vs. Ignorance. As early as 2-years of age, children can accurately acknowledge what they themselves know and do not know – such as on the basis of varying levels of perceptual access (i.e., whether or not they have seen what is inside a container; e.g., Pillow, 1989; Pratt & Bryant, 1990; Ruffman & Olson, 1989; Rohwer et al., 2012, Experiment 1; Tardif et al., 2005), or prior knowledge (i.e., whether they are familiar with the names and meanings of real versus fake words; Maritza & Merriman, 2004). Children can also differentiate between themselves and another person, appropriately affirming that they have knowledge that

someone else does not, and vice versa. For example, in a task developed by Wellman and Liu (2004), children were asked to first look inside a drawer so that they see that there is a toy inside. A doll is then introduced, and children are told that the doll has "never ever seen inside this drawer," and subsequently, asked if the doll knows what is inside the drawer and whether the doll has looked inside it. By the age of 3, children reliably answer these knowledge access questions correctly.

Moreover, when faced with conflicting reports, preschool age children can talk explicitly about how individuals may vary in what they know, reason about what an informant is likely to know, as well as consider how informants' reliability as a source of information is dependent on what knowledge they have. For example, when presented with conflicting claims about a box's contents from an informant who looked inside a box and an informant who did not, 3-year-olds typically trust the claim made by the informant who saw the box's contents (Robinson & Whitcombe, 2003; Whitcombe & Robinson, 2000). Thus, preschool age children's epistemic standards in judging the credibility of the source of new information show the beginnings of an appreciation of judging knowledge claims on the basis of the experiences generated by informants' observed behaviors (e.g., Brosseau-Liard & Birch, 2011; Terrier et al., 2016).

Sources of Knowledge. Knowing about the source of a belief plays a critical role in evaluating how warranted or justified the belief is and how easily it should be discarded. When learning from others' testimony, we are accepting their testimony as evidence of the truth, but not all testimony is equal. When it comes to evaluating testimony, there are a variety of informant-qualities and behaviors that can be used to infer the origins of their belief and, as a consequence, how reliable it is.

The developmental achievement of false belief understanding coincides with corresponding improvements in children's ability to make judgements about informants' knowledge-states beyond knowledge versus ignorance. Specifically, children's epistemic standards show increasing sensitivity to the *process* by which someone learned the information that they transmit to others (i.e., direct observation; Brosseau-Liard & Birch, 2011; Terrier et al., 2016). For example, when presented with two accurate informants, 4- and 5-year-olds (but not 3-year-olds) are more likely to request new information from an informant whose prior accurate testimony was consistently self-generated than from an informant whose prior testimony relied on help from a third party (Einav & Robinson, 2011). Children 4- and 5-years of age can also evaluate the quality of arguments, indicating that looking, testimony from an authoritative source (e.g., teacher), and inference are better reasons for belief than pretense, guessing, and desiring, whereas children 3-years of age struggle more in evaluating the quality of arguments in this way (Koenig, 2012). By age 6, children start to integrate this knowledge with a sense of what kinds of interactions with the world are more or less suitable for learning in certain domains, choosing to 'look' more for knowledge that can be acquired through direct observation (e.g., color) than for that which is impossible to discover through visual perception (e.g., if a person knows French; Fitneva et al., 2013). I

Thus, applying epistemic standards when making judgements of reliability relies on a conceptual understanding of the range of the learning methods possible for acquiring true beliefs in a particular domain. Indeed, the world is inundated with information; information only becomes evidence when considered in relation to a problem. Notably, there is major developmental change between ages 5 and 10 in children's abilities to differentiate between knowledge that can be acquired directly (i.e., through firsthand experience) from knowledge that

requires cultural transmission (e.g., the Earth is round; germs make people sick; Lockhart et al., 2016).

Moreover, there can be variation in the *strength* of the support proffered by different kinds of evidence within the particular problem space. In a sense, the evidential strength offered by informants' testimony depends on how "close" the evidence (from which a belief was generated) is to its epistemic source. Evaluating what counts as evidence and its relative strength, however, is dependent on the domain of knowledge in question.

Domains of Knowledge. While there are many different ways of organizing knowledge, the current dissertation considers how children come to appreciate the distinctions between domains for which the *cultural* world and *natural* world provide the focal interest. Specifically, the current dissertation focuses on children's developing understanding that the cultural and natural domains can be differentiated not only by the range of learning methods possible for acquiring true beliefs, but by the experiences that *better* corroborate a claim based on epistemic standards of evaluation.

The *cultural domain* refers to knowledge that is socially constructed and can only be acquired through social transmission. Language, religious beliefs, cultural customs, and social norms are all within the cultural domain. When evaluating informants' cultural knowledge, the epistemic source is the society of that culture. For example, when the knowledge in question pertains to a language-specific object label, an informant's relation to its epistemic source corresponds to the degree of justification provided for belief in that claim. Testimony from an individual of that culture provides higher justification than from a member of an out-group, with corroborating reports from other members of that culture further justifying the claim.

The *natural domain* refers to knowledge about the physical world, such as information that can be acquired through perception, as well as inferences based on physical evidence. Knowledge about whether something exists or the physical location of an object, as well as entities that are impossible to directly observe (e.g., atoms), fall under this domain. For example, consider the assertion that a particular tree is 7 feet tall; both a testimonial report from an informant with direct experience in measuring its height as well as a testimonial report from an informant who acquired this information second-hand from another individual can be sources of evidence for its truth. A conceptual understanding of the possible pathways to obtaining that knowledge recognizes both as legitimate reasons for accepting that claim. However, in a case where the two testimonial reports offer conflicting claims, the epistemic standard is stronger when the source bases their report on direct experience substantiating it (e.g., measuring the tree's height, in this example), as opposed to more mediated means (e.g., second-hand information about the tree's height, in this example).

The majority of the work on trust and testimony that focuses on the natural world involves knowledge that can be acquired through direct observation (e.g., what is inside of a box, whether a mouse went through a blue door or green door). But in everyday life, we make decisions based on a variety of scientific claims that are not necessarily possible to substantiate via direct observation. Yet, our evaluations still include reasoning about how close evidence to an underlying claim is to its epistemic source, albeit more indirectly, through an understanding that discerning its epistemic source requires expertise. Social information such as labels of expertise become a source of evidence for the reliability of testimonial reports. In turn, the greater extent to which we understand how clusters of different areas of expertise are distributed

in other peoples' minds, the more accurately we can represent the *relations* between converging sources.

Domain-Dependent Distributions of Knowledge

Knowledge propagates through social networks, and yet not all knowledge flows through the social world through the same channels. In fact, direct knowledge does not necessarily require the input of others to be learned. However, certain types of knowledge are arguably more likely than other types to be acquired through interactions with others, and further still, certain types of knowledge are more likely to be acquired through interactions with members of particular social groups.

Past research suggests that adults and children are sensitive to the distinctions between different kinds of knowledge and how widely they are shared (e.g., Cimpian & Scott, 2012; Lockhart et al., 2016; Sóley, 2019; Vélez & Gweon, 2019). For example, around the age of 5, children understand what knowledge can be acquired through firsthand experience (establishing a fact), as well as what knowledge would be easy or what would be difficult, but not impossible, to acquire through firsthand experience (e.g., Lockhart et al., 2016). In addition, consistent with Csibra and Gergely's (2009) theory of natural pedagogy which claims that information presented in a communicative context is more likely to be encoded as both generalizable and culturally shared, 4- and 5-year-olds expect for generic factual knowledge to be shared more widely than non-generic knowledge (Cimpian & Scott, 2012). This is consistent with work from Caza and colleagues, showing that children younger than five struggle to understand that some people (e.g., babies) may not know generic facts about the world (Caza et al., 2016).

Moreover, a growing body of work suggests that from early on, children are sensitive to the boundaries of culturally shared knowledge (e.g., Sóley, 2019). For instance, 5- and 6-year-

olds (and adults) generalize knowledge of songs across people who speak the same language, as opposed to sharing the same gender; however, they do not generalize *preference* for songs across either category (Sóley & Aldan, 2020). Moreover, 3- to 5-year-olds infer that those who share an accent live in the same geographic location, but do not necessarily share personal preferences (e.g., for an activity or a game; Weatherhead et al., 2016). Thus, tracking others' relationships with one another and their broader social network supports our ability to make inferences about what knowledge is likely to be shared and among whom. However, the human social world is comprised of a complex web of social relations and hierarchies; people can belong to many different social categories at the same time. Children must use their knowledge of others' social relations flexibly in determining the significance of different social category markers in social learning.

Nevertheless, each interaction with another person brings new data, and thus, each time a source provides testimony, their social identity and the kind of knowledge required for their assertion to be credible may be consistent or inconsistent with these expectations. In turn, the (in)consistency with expectations of the relations between a testimonial report and its source is likely to influence the degree to which the testimonial report impacts the recipient's belief in the claim. But as emphasized in every statistics course, correlation is not *causation*; if misapplied, such assumptions about the causal power of mere membership of social categories for knowledgeability can lead to dangerous over-weighting of evidential value. So children must be able to keep track of multiple hypotheses and their associated probabilities to selectively learn from reliable sources. To consider this, it is necessary to investigate children's reasoning and evaluation of the reliability of evidence as a process of information aggregation.

Reframing Children's Judgements of Reliability as a Process of Information Aggregation

Four decades ago, psychologists introduced the idea of scientific theory revision as analogous to cognitive development. In keeping with Piagetian constructivist traditions (Piaget & Cook, 1952), the role of prior knowledge in shaping the inferences we draw from new data is central to this account. In brief, the "theory theory" claimed that children construct intuitive theories of the world and alter and revise those theories in light of new evidence (Carey 1985, 1988; Karmiloff-Smith & Inhelder, 1974; Karmiloff-Smith, 1988; Gopnik & Wellman 1992, 1994; Gopnik 1984; Gopnik & Graf, 1988; Keil 1989; Perner 1991; Wellman 1990; Wellman & Gelman, 1992). Much like scientific theories, these intuitive theories are comprised of an ontology of concepts, and a system of (causal) laws that govern how the different concepts interrelate. Intuitive theories enable us to make wide-ranging predictions about what will happen in the future, as well as counterfactual inferences about what could have happened or what would happen if you decided to do something new in the future. Moreover, and crucially, intuitive theories shape our interpretations of new data, influencing our reasoning about what information even counts as evidence.

Both in science and in childhood, information about the probabilistic contingencies between events (i.e., statistical information) play an important role in theory-building (e.g., see Gopnik et al. 2004). Over the last two decades, there has been a great deal of research demonstrating the power of early statistical learning in how children acquire conceptual knowledge of physical, biological, and psychological phenomena (e.g., Gopnik et al., 2004; Gopnik & Wellman, 1994; Schulz & Gopnik, 2004; Wellman & Liu, 2007). The overarching question guiding much of this work has been the following: *How do children learn about the causal structure of the world from patterns of data*? The predominant approach by which this

question has been examined is to show that children's causal learning is best described as a process of rational inference. In the context of children's causal learning, the term "rational" is used to specifically reference the process by which children interpret new evidence (evidence in the form of probabilistic relations)—in light of their current knowledge—in the form of beliefs or "theories" about the world. A large body of work shows that children integrate domain-specific conceptual knowledge into their causal inferences in novel situations (e.g., Denison & Xu, 2010; Schulz et al., 2007; Sobel & Munro, 2009).

Evaluating Evidential Strength from Probabilistic Relations

In the many contexts where someone does not have the expertise to evaluate a claim directly, their evaluation relies on information from external sources. In these cases, the importance of accurately representing the relations between converging sources is paramount. While multiple converging reports can provide stronger evidence than one informant alone, there are many possible routes by which people can come to share beliefs about the world. Critically, the strength of the support offered by converging sources is dependent on how their reports are aggregated. The aim of the sections that follow is to build a bridge from prior work on (a) how children aggregate information over the course of *multiple* events to evaluate the reliability of testimony from a *single* source to (b) how children aggregate information from *multiple* sources for a *single* event: a consensus.

The Influence of Dependencies on Judgments of Reliability

In this section, I review three types of dependency structures. The first type to be discussed is a *chain dependency*, a temporal-based dependency which involves tracking a common thread of influence across a sequence of events. The second dependency discussed is

when we find that multiple converging reports rely on the same *shared evidence*. The final dependency, the main focus of this dissertation, is when individuals share a *common background* that can shape their interpretation of data, and in turn, influence judgments about what can even count as evidence.

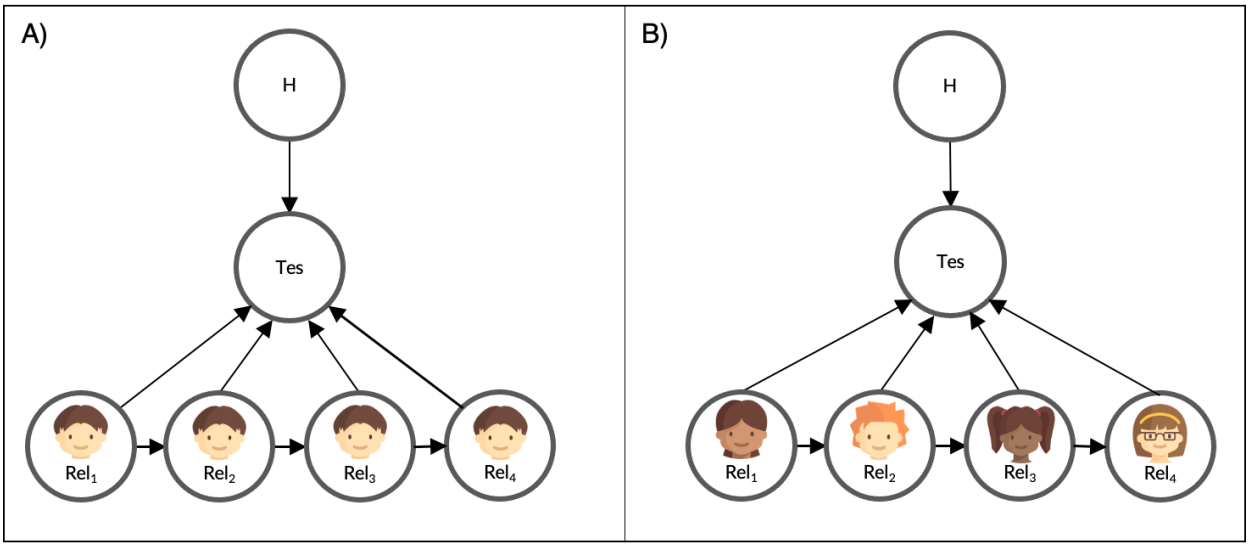
By far, the type of dependency that is most represented in the literature is the chain dependency, although not generally framed in such terms (e.g., Lane et al., 2018; Marble & Boseovski, 2019). It is worth noting that this is necessarily the starting point given that the emergence of a representational Theory of Mind reflects this type of dependency, namely that someone's current knowledge state is dependent on past experience. However, and critically, I will attempt to make the case that these three forms of dependencies—chain dependency, shared evidence, and common background—are interrelated.

Chain Dependencies

Natural Domain. ***Reasoning About a Single Source.*** Butler et al. (2018) presented children between ages 3 and 7 with scenarios in which characters encounter four containers and made factual claims about their contents (e.g., "There are pencils in all the boxes!"). The characters who made those claims first verified them by looking inside all four of the containers, only one of the containers, or none of the containers. Figure 1A shows a graphical representation of this scenario. The variable *H* represents the "true" state of the world, here being the true contents of all four containers. The variable *Tes* represents the testimonial report, which is the claim about the contents of all four containers. The variable $Rel_{1...4}$ represents the chain dependency: the individual reliability of the informant across time, here reflecting the experimental manipulation whereby the behaviors of the characters who made these claims were first observed prior to children's evaluations. The arrows leading from the reliability variables

(as well as the hypothesis variable) specify that the informant's testimony is dependent upon both the "true" contents of the containers and their reliability as a source. Thus, if the informant is reliable, it is assumed that the testimony reported reflects the true state of things.

Figure 1. *Graphical Representation of Chain Dependency Structure*



Note. (A) Chain dependency structure of a single source over the course of multiple events, referenced in discussion of Butler et al. (2018) and Pasquini et al. (2007). (B) Chain dependency structure of multiple sources on one event, referenced in discussion of Hu et al. (2015) and Einav et al. (2018).

Consistent with the predictions of this model, results from Butler at al. (2018) revealed that children across all ages rated verified claims to be more acceptable than insufficiently verified ones. In other words, children across all ages were able to consider the behavior of verifying each of the boxes' contents by the same informant to be the common thread of influence on the validity of the testimony they provided. Importantly, results from these experiments demonstrate a direct relationship between children's explicit understanding that

looking inside the containers would generate the necessary evidence (i.e., perceptual experience) for substantiating a claim about their contents and the extent to which they differentiate between verified and unverified claims. But how can children use both the true state of the world and the reliability of the informant to evaluate their testimony? With respect to the true state of the world, in such a scenario, when the claim at issue relates to an observable reality, children must use their domain-based understanding of the *kind* of experiences that make such an informant reliable (i.e., looking). With respect to the reliability of the informant, they must also understand that only by verifying the contents of *each* of the four boxes would they be able to sufficiently substantiate a claim regarding all boxes' contents.

Recall that perceived reliability is a combination of evaluations of knowledge and intentions, so although not directly examined by Butler et al. (2018), the question remains open as to how children aggregate information about the informant (i.e., behaviors) over the course of multiple events. One possibility is that, with age, children can more easily associate the number of relevant behaviors relative to the total number of behaviors required to make a reliable claim about all boxes' contents. However, another possibility may be that change in performance with age reflects more complex interaction between children recognizing informant's epistemic access (i.e., evaluations of informants' relative expertise) and whether or not they chose to use it (i.e., evaluations of informants' relative trustworthiness).

Reasoning About Multiple Converging Sources. In Hu et al. (2015), children were first shown two boxes and told by an experimenter that each contained a toy (or a snack). Children then watched as individual characters each asserted their opinion about which box contained the better option. One consensus group was comprised of four informants who had each looked inside the boxes prior to asserting their opinions. The other consensus group was comprised of

only one informant who had looked inside the boxes while the other three offered opinions based on hearsay from the informant who offered their opinion before them.

Figure 1B above shows a graphical representation of this chain dependency. The variable *H* represents the "true" state of the world, here being which box contained the "better" option. The variable *Tes* represents consensus members' testimonial reports, which is their converging claim indicating which box contained the better option. The variable $Rel_{1...4}$ represents the chain dependency: the individual reliability of each consensus member across the four timepoints. Thus, the first informant who looked inside the box in the latter consensus group served as the common thread of influence across the sequence of events, here being the three subsequent opinions offered.

Hu et al.'s (2015) results revealed that 4-year-olds can differentiate between two consensus groups equal in size based on *how* each group's respective members came to believe what they agreed upon. When asked to choose which of the two boxes had the better option, children chose the box endorsed by the consensus group comprised of informants whose beliefs were *independently* credible, that is, the evidence required to substantiate their reports was generated separately from one another. Similar to Butler et al. (2018), a claim by a consensus corroborated by four informants is best supported by all four informants checking each of the boxes prior to making their claim.

Consistent with the previously mentioned possibility regarding Butler et al. (2018), children's evaluations of consensus accuracy may reflect their ability to merely associate the number of relevant behaviors (i.e., looking inside both boxes) relative to the total number of behaviors possible across sources. However, while the chain dependency provides a similar structure to Butler et al. (2018), it no longer relies on a single informant. In turn, unlike in Butler

et al. (2018), the subsequent testimony dependent on that first claim, while based on hearsay, is not void of evidence. Rather, it reduces the distinct pieces of evidence that would have been possible given each informant's own personal epistemic access. Hu et al. (2015) conducted a follow up study using the same paradigm but the consensus group with the chain dependency now had more members than that of the independent consensus. Adding individuals to the chain dependency allows for a clearer test of whether 4-year-olds' preference for an independent consensus reflects understanding that convergence between sources is *better* supported by as many possible unique sources of evidence closest to its epistemic source (direct observation).

Results from Hu et al.'s (2015) follow-up study revealed that 4-year-olds now lacked a preference, endorsing the consensus with the chain dependency and the independent consensus at equal levels. This suggests that the results from Hu et al.'s first study may not reflect 4-year-olds' understanding that the other consensus members' reliance on the first informant's report (i.e., consensus with a chain dependency) reduces the evidential support, as compared to converging reports that rely on unique sources of evidence. Thus, once again the question remains open as to how children aggregate information across consensus members to evaluate the reliability of what it is that they agree upon.

In line with the perspective that the perceived reliability of a source is a combination of evaluations of their expertise and trustworthiness, 4-year-olds' *overvaluing* of consensus members' reliance on second-hand information perhaps reflects an interpretation of their endorsements as signifying the first informant's social standing or the social value that they possess. This interpretation can lead to rationalizing the decision to assign greater weight to an opinion made by one informant than opinions made by multiple informants, on the basis of their higher social standing. Thus, younger children consider the relations between consensus

members for evaluating the reliability of their converging viewpoint. Children's lack of understanding that it is possible for people to have different interpretations of the same event leaves the choice of which perspective to trust as more of an endorsement of the individual's social influence, rather than a reflection of an epistemic standard.

Cultural Domain. ***Reasoning About a Single Source.*** In two studies, Pasquini et al. (2007) assessed 3- and 4-year-olds' sensitivity to the previous accuracy of informants, manipulating the relative frequency of their inaccuracies as follows: 100% vs. 0%, 100% vs. 25%, 75% vs. 0% and 75% vs. 25%. Over the course of four events, children viewed two informants label familiar objects (e.g., spoon; shoe). Children subsequently observed the same informants provide conflicting object labels for *unfamiliar* objects and were asked to identify the correct label for each object.

The graphical representation in Figure 1A above shares the same structure as Butler et al. (2018). The variable *H* represents the "true" label of the unfamiliar object on the final test trial. The variable *Tes* represents the testimonial report, which is the claim about the unfamiliar object label on the test trial. The variable $Rel_{1...4}$ represents the chain dependency: the individual reliability of the informants across time, here reflecting each informants' individual history of accuracy in labeling familiar objects in the experimental manipulation. Here too, the arrows leading from the reliability variables (as well as the hypothesis variable) specify that the informant's testimony is dependent upon both the "true" label of the unfamiliar object and their reliability as a source. Thus, if the informant is reliable, it is assumed that the testimony reported reflects the true state of things.

Results from Pasquini et al. (2007) indicate that 4-year-olds successfully tracked the two informants' relative frequency of errors across all conditions and used this information to guide

their trust on the test trial. In contrast, the 3-year-olds appeared to simply be monitoring informants for any display of inaccuracy, choosing to endorse the labels provided only by an informant with a history of 100% accuracy. The difference in performance between 3- and 4-year-olds can be interpreted with respect to variations in how they aggregated information across the chain dependency (see Shafto et al. 2012).

Recall that evaluations of perceived reliability are a combination of inferences of informant knowledge and trustworthiness. An important distinction between 3 and 4-year-old children's performance parallels developmental change in false-belief understanding. Three-year-olds' evaluations perhaps reflect their aggregation of inferences of informants' knowledgeability, based on their accuracy over the course of multiple events. Any informant that provides inaccurate testimony for one event is then deemed equally fallible as informants who provide inaccurate testimony for multiple events. Whereas four-year-olds' evaluations perhaps reflect a more nuanced and complex interaction of inferences related to both informant knowledgeability and their helpful intentions to provide accurate information. Aggregating information across both of these dimensions lead to better justification of trust in informants who are fallible but have good intentions.

Reasoning About Multiple Converging Sources. Einav (2018) investigated the influence of a chain dependency on children's evaluation of consensus reliability. Participants between ages 5 and 9, as well as adults for comparison, were introduced to a 'faraway country' of Markovia. The experimenter then presented a book about Markovia, indicating gaps in its content. Participants were told that it would be their job to fill in the book's gaps with the correct pictures, and that they would have access to informants who were previously taught a lesson about the unfamiliar country. Participants were shown short vignettes in which two groups of

informants were each asked a question by an off-screen speaker. The informants always non-verbally responded to the question by holding up a picture card in the air.

One consensus group comprised of individuals who each looked straight ahead when they responded (i.e., informants could not see each other's answers). For the second consensus group, when the first informant held up their chosen picture card, the other two informants leaned over and looked at the first informant's picture card (i.e., chain dependency). Thus, Einav (2018) shares the same event structure as Hu et al. (2015); Figure 1B serves as a graphical representation of this chain dependence. The variable *H* represents the "true" state of the world, here being knowledge about the unfamiliar country (e.g., the type of housing in Markovia). The variable *Tes* represents consensus members' testimonial reports, which is their converging claims regarding the unfamiliar country Markovia. The variable $Rel_{1...4}$ represents the chain dependency: the individual reliability of each consensus member across the three timepoints.

Einav's (2018) findings suggest that there is a developmental shift in children's reasoning about the influence of the chain dependency on consensus reliability. Consistent with the adults, 8- and 9-year-olds were more likely to accept the claims provided by the group whose responses were independent and justified their selections referring to their independence. Five- and 6-year-olds were at chance, lacking a preference for either consensus group. While 5-year-olds were not systematic when explicitly asked to justify their choice, 6-year-olds on the other hand, were almost equally split in terms of justifying their endorsement for the independent group by referencing independence and for the group with the chain dependency by referencing the fact that informants looked at another's answer.

While the patterns of results are consistent with the research discussed earlier, Einav (2018) is the only study that asked children to justify their choices. Younger children's tendency

to refer to the fact that informants looked at another's answer in justifying their preference for the consensus with the chain dependency, suggests that their lack of preference for the independent consensus cannot be due to their failure to recognize the dependency or even its influence on consensus members' reports. Consistent with the interpretation of Hu et al.'s (2015) second study, perhaps, here too, younger children's preferences reflect that they interpret subsequent informants' endorsements of the first informant's answer as a signal of the first informant's social standing or the social value. In interpreting the chain dependency as a signal of the first informant's reliability as a source—comprising both evaluations of expertise and trustworthiness—the decision to assign greater weight to an opinion made by one informant with higher social standing can be rationalized as justified.

What accounts for the qualitative change between how younger and older children evaluate the impact of a chain dependency on the evidential support offered by a consensus? In order to recognize how a chain dependency results in redundant information (reducing evidential support), children must have an appreciation for *interpretive diversity*. That is, because beliefs are a product of the mind's interpretation of experience, it is possible that individuals can have different interpretations of the same information. In Einav (2018), children were told that all informants were taught the same lesson about Markovia. Thus, without an understanding that it is possible for those informants to have had different interpretations of that lesson, it is unlikely for children to recognize how evidential strength is reduced when those answers rely on a single informant's interpretation (i.e., the chain dependency).

Nevertheless, recognizing that evidential support is reduced by chain dependencies requires understandings beyond mere appreciation for interpretive diversity. The possibility that younger children interpret the chain dependency as a signal of the first informant's higher social

standing can still be rationalized based on the perspective of one informant being weighed as more valuable than others due to their social standing. Thus, children must have an ability to recognize that while two individuals can have diverging positions on the same issue, one position can have more merit ("can be more right") to the extent to which it is supported by *unique* sources of evidence.

In the cultural domain, knowledge about the culture of a faraway country such as in Einav (2018) requires the input of others. Children were told that members of both consensus groups were taught the same lesson about Markovia, and thus that lesson served as the epistemic source against which children evaluated the relative evidential strength offered by both consensus groups. Knowledge about a culture can only be learned through social transmission, making information passed along through chain dependencies not only a valid source of evidence for acquiring true beliefs about that culture but the *only* source of evidence from which cultural knowledge can be acquired. Both the lesson about Markovia itself, as well as the consensus groups' converging reports, represent evidence for given claims about Markovia's culture that can be traced back to a chain dependency. Only by age 8 did children demonstrate an understanding that the extent to which each consensus group's members' converging reports represent *distinct* chain dependencies determines the evidential value of each consensus group's converging reports.

Interim Summary

Thus far, I have presented studies that use a single source paradigm, coupling each with an analogous study in which children evaluate claims from multiple converging sources to analyze chain dependencies. The purpose of coupling findings from single informant paradigms with those of multiple informant paradigms was to demonstrate that a testimonial report from a

single source can be characterized as a chain dependency. An informant's past experiences shapes their interpretation of new experiences, and as a consequence, what they come to believe and report to others. It follows that when informants provide converging reports based on their own direct experience (e.g., independent consensus groups in Hu et al., 2015 & Einav, 2018), each report represents a *distinct* chain dependency. Given this, overlap across chain dependencies that underlie converging reports leads to redundant information (reducing evidential value).

However, in most circumstances it is simply not feasible to trace the common thread of influence across the sequence of events that led converging sources to believe what they agree upon. Doing so requires access to informants' unique social histories and searching for potential points in time for which the chain dependencies may have overlapped. In daily life, this sort of information is particularly scarce, and even if accessible, such an endeavor is extremely cognitively taxing. Moreover, there are many circumstances in which overlap across chain dependencies can be inferred without information regarding the timeline of events that led each source to believe what they report to others.

As an example, consider two students who provide converging answers to a math problem. Each student represents a chain dependency in that their reports are dependent upon their unique experiences. Likewise, the evidential support for the accuracy of that answer is also dependent on their unique experiences. However, if you learn that both students copied the answer of another classmate, the point in time for which each student did so can be represented as a point of overlap in the chain dependencies that underlie their converging answers. Critically, the timepoint of each event where the students copied their classmate's answer matters less so than the fact that their converging reports relied upon the *same* classmate's answer to the math

problem. Said differently, whether the two students copied their classmate's answer together during morning announcements, or one student copied the answer before school and the other during lunchtime, does not change the fact that the evidential support offered by the two students' converging reports are dependent upon the same classmate's answer. What is important to consider in evaluating the evidential strength of the two students' converging reports for the accuracy of their answer is that they relied on *shared evidence*.

In the following section, I review a set of representative studies in the developmental literature that focus on how direct information about converging reports' reliance on *shared evidence* influences children's judgments of reliability. The aim of this section is to highlight the parallels between developmental trends in the influence of *chain dependencies* and *shared evidence* on children's judgments of consensus reliability. Importantly, irrespective of the domain, the extent to which converging reports may rely on shared evidence corresponds to the amount of evidential support offered by those reports. Therefore, while the previous section was organized by natural and cultural domains, the studies reviewed in the following section focus on when in development children come to understand that converging reports that rely on shared evidence in any domain lead to redundant information.

Shared Evidence

There are few studies that directly examine when in development children come to appreciate that the evidential value of converging reports is reduced when those reports rely on *shared evidence*. However, in the studies relevant to the issue, children are provided with *direct* information about the evidential source of informants' reports. Recall that as outlined in the chain dependency section, while children 4 years of age can recognize verified claims as more acceptable than insufficiently verified ones in single informant paradigms (Butler et al. 2018),

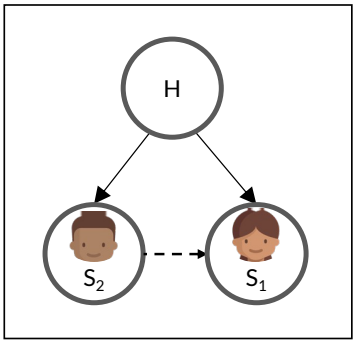
they fail to recognize that a consensus originating from more *distinct* sources of direct evidence is *better* evidence for a claim than a consensus based on a single item of direct evidence and subsequent hearsay (Hu et al., 2014). While both firsthand experience and others' testimony can lead to knowledge about the natural world, when the knowledge in question is possible to acquire (in principle) directly via observation, the reliability of related knowledge claims depends on the relation between the evidence and its epistemic source. Specifically, the origins of the kinds of information that count as evidence can be traced back to its source, potentially along a multitude of chains of dependency before arriving at the firsthand experience of an individual. The essential foundation for recognizing that converging reports based on one informant's direct experience and subsequent hearsay, is the ability to understand that direct experience offers stronger evidential support for a claim than second-hand information, an understanding that continues to develop through middle childhood. Several studies suggest an age-graded decrease for preferring second-hand (i.e., socially mediated), as opposed to first-hand (i.e., direct observation) testimony (e.g., Fitneva, 2008; Matsui et al., 2006). For example, Lane et al. (2018) presented children ages 3 to 8 with individual informants who made claims about improbable (e.g., someone drinking onion juice) and impossible events (e.g., someone turning applesauce into an apple), and told the children that their help was needed to figure out whether different things can happen in real life. Given that children between 4 and 8 years of age increasingly believe that it is possible for improbable events to occur (e.g., Shtulman & Carey, 2007), Lane et al. (2018) hypothesized that children's beliefs about improbable (as opposed to impossible) events may be most sensitive to the influence of other inputs, such as others' testimony.

In Lane et al.'s (2018) study, children watched a series of videos of informants making claims. In two videos, an informant provided first-hand claims (e.g., "I saw someone drink onion

juice"), and in the two remaining videos, an informant made claims based on second-hand information (e.g., "Someone told me they drank onion juice"). After watching each video, children were asked: (a) whether the event could in fact happen (e.g., "So, what do you think? Could a person drink onion juice in real life, or not?"), and (b) how sure they were about their answer (e.g., "Okay, you think that a person [could/could not] drink onion juice in real life. Are you very, very sure or just a little sure?").

Figure 2 shows the graphical representation of the shared evidence that underlies the second-hand informant's claim in this context. The variable *H* represents the "true" state of the world, here being whether the event could occur in real life. The S_1 and S_2 variables represent the experimental manipulation: the source's report based on second-hand information. Here, S_1 represents the source's report based primarily on the claim of someone else (S_2), rather than their own direct observation of the improbable event. Findings from Lane et al. (2018) revealed that there was a significant age-graded decrease in belief that improbable events could in fact occur in real life after children heard second-hand claims, but not after hearing first-hand claims. Specifically, 4-year-olds reported *greater* certainty in the possibility of improbable events after hearing second-hand claims, whereas by the age of 8, children reported *less* certainty. Children's certainty that impossible events could in fact occur in real life, on the other hand, remained unchanged across age groups.

Figure 2. *Graphical Representation Shared Evidence Dependency Structure*



Note. Shared evidence dependency structure, referenced in discussion of Lane et al. (2018) and Aboody et al. (2022).

In another set of studies, Aboody et al. (2022) systematically varied the number of informants with the number of direct sources of evidence they cite across three experiments with children ages 4 to 7, using the same paradigm. Children were told that there was one blue and one green door leading to a classroom. This classroom had a pet hamster that ran away that morning and the teacher needed to look for it. The teacher wanted to know whether the hamster went through the blue or the green door. In the first study, two students appeared on a screen and the experimenter said that one student claimed the hamster ran through the blue door while the other student claimed the hamster ran through the green door. Children were told that the teacher asked each student, "Can you tell me why you said the hamster ran through the blue door?" One student answered, "Because Mary said she saw it there, because Kayla said she saw it there, and because Sarah said she saw it there," thus citing three distinct sources with direct evidence. The other student answered, "Because Annie said she saw it there," thus citing one source with direct evidence. The experimenter then asked the test question, "Can you tell me, where is the hamster? Through the blue door or through the green door?"

In the second experiment, with the same paradigm, six students appeared on the screen, with three students claiming that the hamster ran through the blue door and three students claiming the green door. Children were told that three students in one group responded to the teacher's question asking why they made that claim and cited three distinct sources with direct evidence. The other three students responded citing a single source with direct evidence. That is, the converging reports made by the second group of three students relied upon *shared evidence*. Thus, Figure 2 can be expanded to reflect this dynamic as well. In the final experiment, one student appeared on each side of the screen. Children were told that one student responded to the teacher's question by citing three distinct sources with direct evidence, and the other student responded citing a single source with direct evidence.

Aboody et al.'s (2022) results revealed that only by age 6 do children prefer a consensus group with fewer informants who cite *multiple* sources of direct evidence over a consensus group with a larger number of informants who cite a *single* source of direct evidence. Importantly, the third experiment here parallels the second experiment in Hu et al. (2015), corroborating the result that 4-year-olds fail to recognize that the evidential value of converging reports is strongest when each report reflects a *distinct* source of evidence closest to its epistemic source (direct observation). Thus, in multiple informant paradigms, there is qualitative change between early and middle childhood in how children account for the impact of both a *chain dependency* and a reliance on *shared evidence* on the support offered by a consensus for a claim. Notably, this qualitative change in children's ability to consider how dependencies can lead to redundant information (thus reducing evidential support) appears to parallel the developmental shift in children's appreciation for interpretive diversity (i.e., iToM; See Carpendale & Chandler, 1996, 1998; Lalonde & Chandler, 2002; Ross et al., 2005).

In the studies reviewed thus far, children were presented with information that *directly* bears on the experiences (and thus the evidence) that underlie agreement between consensus members. However, much like how it is simply infeasible to trace the chain dependencies which led converging sources to their origins, so too is mapping out the sources of evidence that individual consensus members relied upon when forming their converging views. Further, while a reliance on shared evidence leads to redundant information, redundancy in and of itself does not undermine the truth value of a given claim. Building on the expertise of others *in any domain* inherently involves relying on shared evidence that accumulates over time. Science, as an enterprise, is able to advance by building knowledge while relying on shared evidence within and across disciplines.

Moreover, in everyday life it is a common feature of real-world testimony that a recipient does not know exactly *how* individual sources come to believe what they believe and report to others—rather, the only source of evidence available may be specific patterns of variation and covariation between the beliefs people hold and with whom they affiliate. The interrelation between any given society's social networks can represent overlapping channels of information, shaping the perspectives of its members. Overlap in the flow of information through these networks reflects a *common background* through which individuals interpret their experience. In the section that follows, I discuss the final dependency, when individuals share a *common background.* Relevant findings from the developmental literature will be presented, including evidence from preliminary studies which shape the methods and aims of the current dissertation.

Common Backgrounds

Adults and young children alike make use of social group concepts to infer regularities regarding familiar categories such as gender and race, as well as novel groups to which they

themselves do not belong (Fiske et al., 2002; Gelman et al., 2010; Liben et al., 2002; Roberts & Gelman, 2015; Shutts et al., 2013). Many studies have shown that children's epistemic standards include cultural group membership. For instance, 4- to 5-year-olds selectively endorse information from native-accented speakers (Kinzler et al., 2011). Further, children ages 6 to 9 prefer to learn a novel cultural practice from a cultural group member who learned it from another member of their culture over a member of the child's own in-group ("an American informant") who learned this practice from a book (Marble & Boseovski, 2019). This is consistent with previous findings that in middle childhood, children display a developing ability to differentiate between the kind of learning experiences relevant for learning in a given domain and the relative evidential strength of the information they generate.

Children can also distinguish experts based on their domain of expertise, preferring one expert over another depending on the problem at hand (Koenig & Jaswal, 2011; Sobel & Corriveau, 2010). By 8 years of age, children recognize that an individual's social relationships can influence their judgments in subjective contexts (e.g., judging a baking contest; Mills & Keil, 2008). Therefore, depending on the claim at issue, certain patterns of consensus provide stronger evidence of reliability than others.

Given that the human social world is comprised of a complex web of social relations, the question of whether relations between consensus members exist is not necessarily most important. Rather it is whether their overlapping networks reflect experiences that lead to knowledge and whether the origins of consensus members' reports reflect the range of *distinct*, or *diverse* pathways possible for knowledge acquisition. This means that how we represent not only our relationships with others, but also others relationships and the social networks that those relationships comprise, can be used to infer dependencies. On its own, social relations between

converging views can be a source of evidence based on the perspectives represented by individual consensus members. The finding that children's abstract representations of social categories include expectations that categories constrain behavior (and thus experiences) is well-documented, for both familiar social categories such as gender and novel groups to which they themselves do not belong (e.g., Blakemore, 2003; Gelman et al., 2010; Liben et al., 2002; Rhodes & Chalik, 2013; Roberts et al., 2017; Shutts et al., 2010, 2013). Thus, a key question is how children might use different patterns of consensus to infer that a certain position is correct based on the domain of knowledge in question.

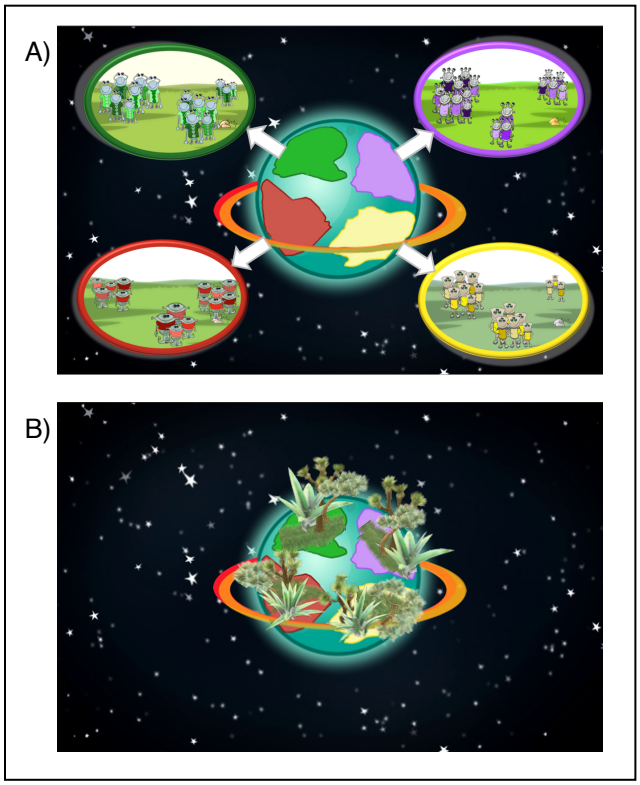
Preliminary Studies. Two preliminary studies (Levush et al., in prep) form the basis of the current dissertation. The aim of the preliminary studies was to investigate 5- and 6-year-olds' (Study 1A), as well as 7-8-year-olds', and adults as a comparison, (Study 1B) capacity to use the social composition of consensus groups as a source of evidence for making judgments of consensus reliability. In both studies, participants were introduced to a novel planet named Planet Kazar and four kinds of aliens that live on Planet Kazar ("Flurps," "Zarpies," "Blarks," and "Gorps"). There were several reasons for choosing to use a novel planet and social groups. First, the use of novel categories ensures that children's responses are not influenced by pre-existing views of familiar social categories. Second, past research using novel categories to examine children's abstract beliefs about social categories has consistently demonstrated that children selectively attribute shared cultural knowledge on the basis of group membership (e.g., Karadağ & Sóley, 2022; Marble & Boseovski, 2019; Roberts et al. 2017; Sóley, 2019).

Each of the four alien social groups were introduced using a noun label (e.g., Flurps) and an accompanying visual of members of the respective alien group. Group membership was also visually marked by shared physical features (e.g., skin color; number, shape, and location of

eyes). Moreover, and critically, social group membership correlated with geographic location, with each alien group living on one of the four continents on Planet Kazar. Each continent appeared on the screen in the same color as the clothing color worn by its respective alien group inhabitants as the experimenter narrated, “Look! the [Gorps] all live together here (continent appears). See, this is where the [Gorps] live and do things that Gorps do!”

Participants were told that there are weird things that happen on Planet Kazar and that they will be learning all about them in the activity—specifically, weird things that “aliens *do* where they live,” (Cultural Phenomena; see Fig. 3a) and weird things that “happen in *nature* everywhere,” (Natural Phenomena; see Fig. 3b) on Planet Kazar—and their task is to figure out why it is that those things happen. The experimenter explained that because no one on Earth has ever seen or heard about Planet Kazar, their job was to make sure that they learn the *right* things so that they would be able to teach everyone on Earth all about it. After learning about a given weird thing, participants were presented with two groups of aliens, each of whom collectively thinks different things about *why* the given weird thing happens, and asked to choose which of the two groups was the *best* group to ask for an explanation.

Figure 3. *Introduction Stimuli for Cultural & Natural Phenomena Trials*

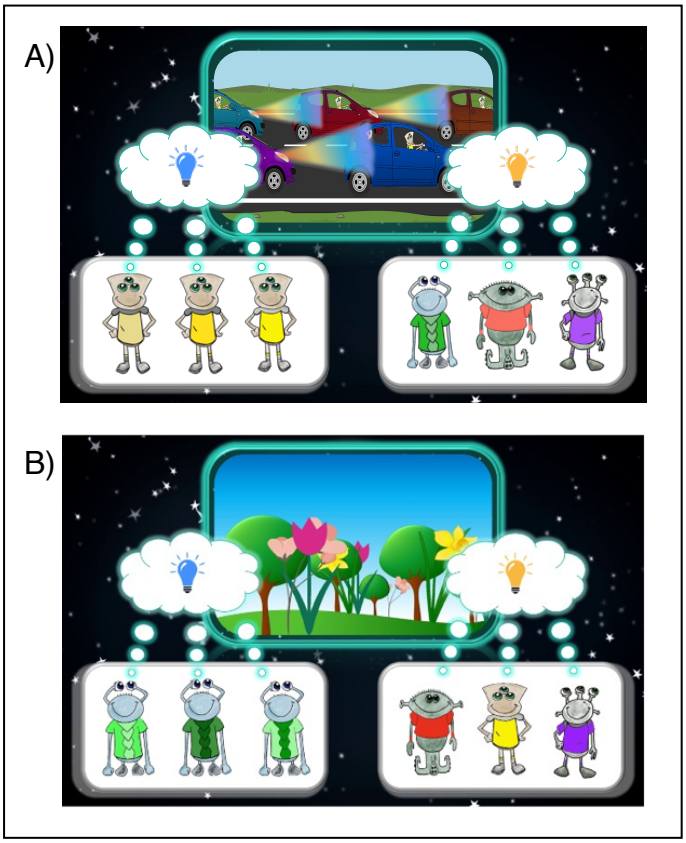


Note. Introduction stimuli for experimenter narration of task instructions. (A) Cultural phenomena referencing weird things that "aliens do where they live on Planet Kazar." (B) Natural Phenomena referencing weird things that "happen in nature everywhere on Planet Kazar."

There were two trial blocks, a Cultural Phenomena Trial Block and a Natural Phenomena Trial Block, each comprised of four trials, with each trial including one "weird" thing that happens on Planet Kazar. The two consensus groups presented across all trials included one homogeneous consensus group, comprised of three aliens who all *share* group membership (e.g., Zarpie, Zarpie, Zarpie), and one diverse consensus group, comprised of three aliens each with *distinct* group membership (e.g., Blark, Flurp, Gorp). On Cultural Phenomena trials, participants learned about group-based regularities of certain behaviors or practices, such as, "Where the Zarpies live on Planet Kazar, they drive cars with rainbow lights!" (See Figure 4A). When

learning about the cultural phenomena, the homogenous consensus was comprised of aliens who all shared membership of the same alien group described in the test item. On Natural Phenomena trials, participants learned about natural phenomena that occurs everywhere on Planet Kazar such as, “Everywhere on Planet Kazar, the flowers grow to be as tall as the trees!” (See Figure 4B).

Figure 4. *Example Items from Cultural Phenomena & Natural Phenomena Trial Blocks*



Note. (A) Example item from cultural phenomena trial block, illustrating same group members all driving cars with rainbow lights. (B) Example item from natural phenomena trial block, illustrating that the flowers grow to be as tall as the trees.

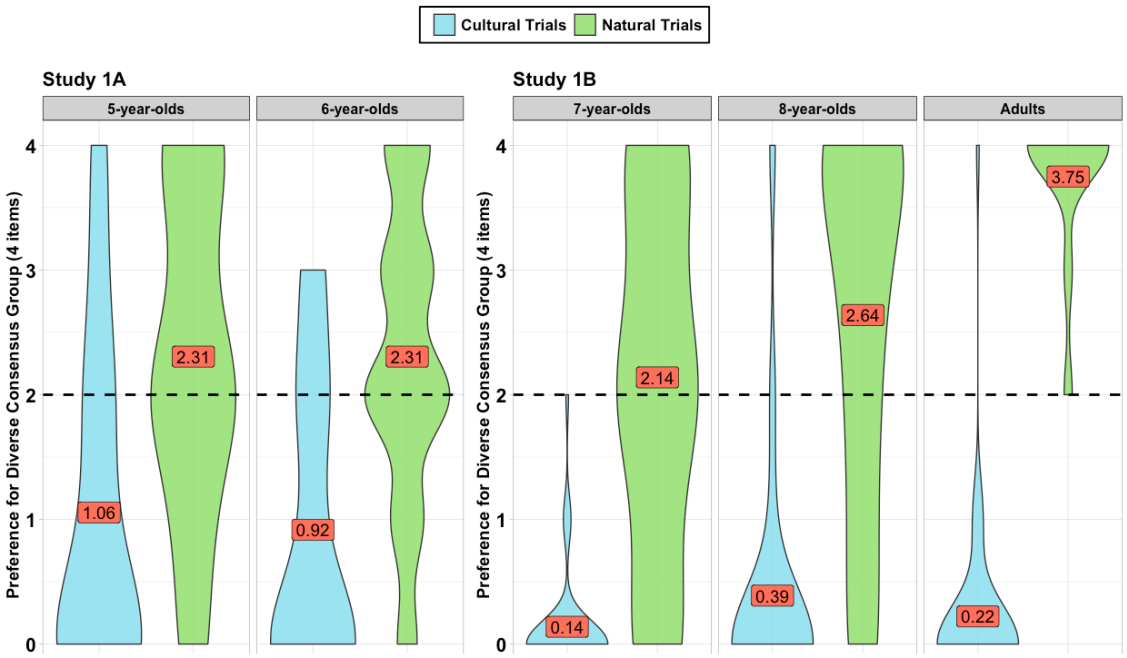
Importantly, participants made their decisions about which consensus group to ask without an ability to evaluate the contents of the consensus members' beliefs. Children could only base their decisions on the domain of the weird thing (natural vs. cultural phenomena) and the different patterns of relations exhibited by each consensus group (homogenous vs. diverse group membership). As mentioned earlier, by age 5, children can appreciate not only that there are different types of knowledge, but that some are culturally specific and thus more likely to be shared among members of that culture (e.g., Sóley, 2019). Thus, for cultural phenomena, it was expected across all ages for children to prefer homogeneity among consensus members, opting to learn from members of the group in which that phenomenon is observed.

As for the natural phenomena, it remained less clear at what point in development children might come to recognize the value of diversity among consensus members and therefore was the main focus of investigation. However, findings from the literature on children's inductive reasoning, using various methodological approaches, suggest that children's understanding of the value of diverse evidence for making generalizations about broader kinds (e.g., category of dogs) develops through middle childhood (Gutheil & Gelman, 1997; Li, et al., 2009; Rhodes & Brickman, 2010; Rhodes et al., 2008; Lopez et al., 1992). Thus, the main hypothesis was that with age, children would increasingly align their preferences for diverse and homogeneous consensus patterns with natural and cultural domains, respectively.

Figure 5 shows both studies' results. In Study 1A, both the 5- and 6-year-old age groups were more likely to select the diverse, as opposed to the homogeneous consensus group for the Natural trials than they were for the Cultural trials. However, an analysis of children's consensus group choices within each trial type revealed that while the 5- and 6-year-olds selected the predicted (homogeneous) consensus group more often than expected by chance for Cultural

trials, they lacked a preference on the Natural trials. In Study 1B, all children (7- and 8-year-olds) and adult participants were also overall more likely to select the diverse (as opposed to homogeneous) consensus group for the Natural Trials than they were for the Cultural Trials. Follow-up analyses within each age group among the child participants revealed differences in performance. Specifically, the 7-year-olds selected the predicted (homogeneous) consensus group for the cultural trials at above chance levels but showed no preference for either group on the natural trials; whereas the 8-year-olds preferred the predicted consensus groups more often than expected by chance for both trial-types.

Figure 5. *Results from Preliminary Studies*



Note. Study 1A-1B: Preliminary studies results. Violin plots showing the relative preference for diverse consensus group across four trials on the *Cultural Phenomena* trial block (blue) and *Natural Phenomena* trial block (green). Higher values indicate stronger preference for diverse consensus group. Red labels show means; black dotted line indicates chance performance.

It is not until middle childhood that children come to recognize the merits of different patterns of consensus for natural and cultural domains of knowledge (Levush et al., in prep). The findings from the preliminary studies that form the basis of this dissertation are consistent with the research reviewed in this chapter. Taken together, it suggests that conceptual understanding of the *relative evidential strength* of the information generated by different methods of knowing perhaps underlies the qualitative change in how children account for relations between consensus members' experiences in their judgments of reliability.

Critically, the strength of the support offered by converging sources is dependent on how their reports are aggregated. Recall that the epistemic value of evidence depends on how "close" evidence (from which a belief was generated) is to its epistemic source. In the preliminary studies, children had no direct information bearing on the evidence underlying the belief shared by consensus members. Beyond mere social group membership, perhaps the *proximity* between consensus members' geographic locations informed children's evaluations. From early on, children can reason about the implications of individuals' direct interactions with the world on the beliefs that they form (Aboody et al., 2022; Brosseau-Liard & Birch, 2011; Nurmsoo & Robinson, 2009; Robinson et al., 2011). What remains less clear is how children come to appreciate the role of social group membership in shaping perspectives beyond their immediate physical environments. Further, the explanatory process underlying children's preferences for different patterns of consensus also remains an open line of investigation. These questions are central to the current dissertation.

Chapter 3: Methods

Sample

Data was collected from 144 children evenly distributed across three age groups – 48 7-year-olds (*M*= 7.50, *SD*= 0.27, 24 females), 48 8-year-olds (*M*= 8.47*, SD*= 0.31, 24 females), and 48 9-year-olds (*M*= 9.42*, SD*= 0.30, 24 females). An additional 12 children were tested but were excluded from final analyses due to experimenter error or technical difficulties [*n* = 6], parent interference [*n* = 5], and insufficient attention [*n* = 1].

Child participants were recruited through social media advertisements for families interested in participating in developmental studies. Participation was limited to children living in the United States. Appendix A shows a summary of child participants' general geographic locations. Parents/guardians received an email invitation for their child to participate which included a direct link to scheduling a virtual appointment with a trained experimenter. Informed consent from parents/guardians as well as verbal assent from child participants was obtained prior to beginning the study activity, and all participants received a $5 gift card for their participation. Prior to scheduling an appointment, parents completed a demographic survey (items listed in Appendix B). Appendix C shows descriptive statistics for these demographic survey variables.

An additional 48 adults (*M*= 37.65*, range*= [19.0, 61.0], 21 females) were recruited through the Prolific research platform as a comparison group. Participation was limited to adults living in the United States. Appendix D shows a summary of adult participants' general geographic locations. Demographic information was obtained through the Prolific platform. Appendix E in the appendix shows descriptive statistics for these demographic survey variables.

Racial Diversity Index: A snapshot of children's school contexts

The school information that parents/guardians provided in the demographic survey was used to obtain school-level demographics. A Racial Diversity Index (RDI) score, adapted from the state of California Ethnic Diversity Index (California Department of Education, 2020), was computed based on this information to provide descriptive information about children's school contexts. The RDI is a measure of the level of diversity or variation in a school's student population across different racial groups. A higher index indicates a more even distribution among the student body. For instance, a school with an equal number of students across all eight racial categories would have an Index of 100, whereas a school with all students belonging to the same race would have an index of 0.

It is important to note that schools with the same RDI score may have very different racial compositions. For example, a school with a 50/50 split of Latinx and White students would have the same index as a school with equal numbers of Asian and Black students, as both schools have two evenly divided racial groups. Thus, the index does not give priority to any specific racial group. Additionally, this Index focuses on demographic diversity and does not encompass other important characteristics such as sense of belonging and influence of systemic racism, among myriad other factors that shape children's school contexts and experiences. Nonetheless, the RDI can provide valuable information about a school's student population in terms of the number of racial groups represented on campus and in the community, which may not be immediately apparent without the index.

Of the 135 parent/guardian reports of school information for their child in the demographic survey, 66.7% attend a public school (n =96), 4.2% attend a public charter school (n=6), 9.7% attend a private school with religious affiliation (n=14), 2.8% attend a private school

with no religious affiliation, and 10.4% are homeschooled (*n*=15). An RDI score was calculated for the 112 of the reported schools in parent/guardian survey responses. School demographic data was obtained from the National Center for Educational Statistics (*n*=109), and state-government/district level sources (*n*=3). Overall, the mean index score was 39.26 with a range of 0.3 to 69.4 (See Appendix F for boxplots of the distribution of RDI scores within each age group). As a point of reference, as of April 2020, the state-wide measure in California indicates that the highest index score for the state is 76 (California Department of Education, 2020).

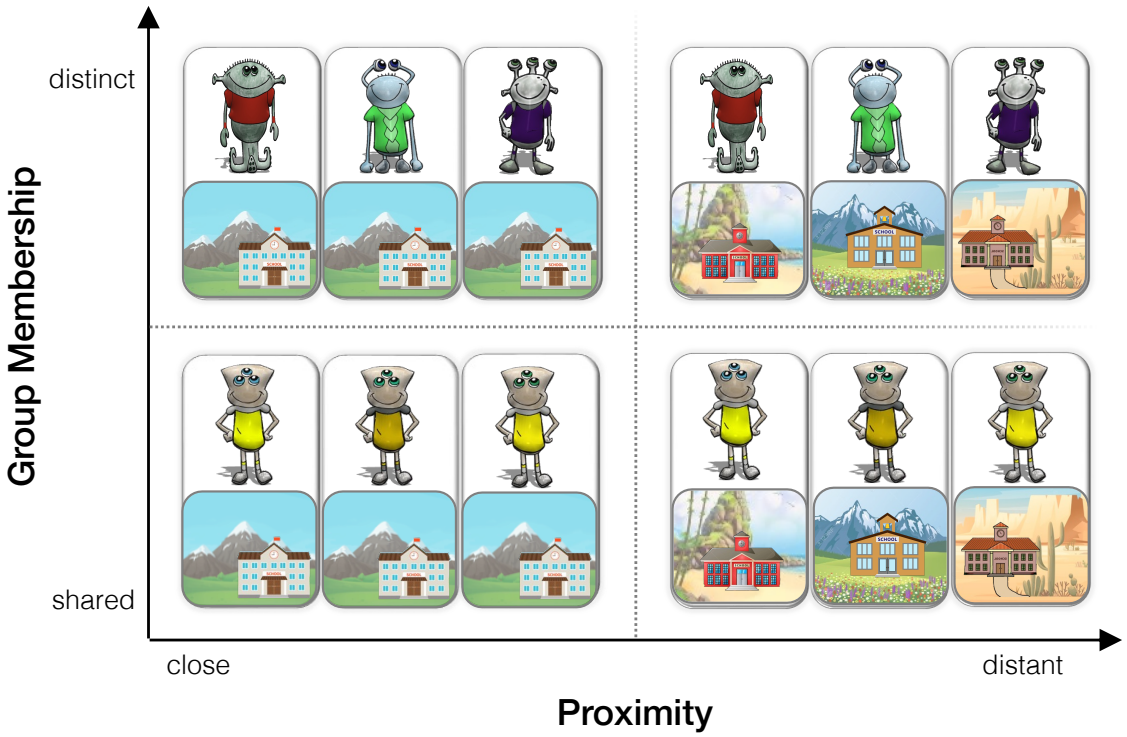
Study Overview

The current study consisted of 2 phases: a story introduction and a test phase. The story introduction involved presenting a novel planet, the four kinds of aliens that live there ("Flurps," "Zarpies," "Blarks," and "Gorps"), and explaining the structure of the task. Generic noun labels (e.g., Flurps) were used to identify social group membership. Social group membership was also perceptually marked by shared physical features, as well as clothing color worn by each groups' respective members. Participants were told that much like children on Earth, everywhere on the alien planet there are members of each alien group attending school. Thus, both the dimensions of social group membership (*shared* vs. *distinct*) and proximity (*close* vs. *distant*) relational properties were manipulated with the goal of examining how children represent, reason, and integrate each of these cues as a unique source of influence on the experiences underlying agreement between consensus members.

The test phase consisted of 3 trials, all in which participants were asked which of 2 consensus groups, one with *shared* group membership and one with *distinct* group membership, is the better group to ask to help them learn why a particular natural phenomenon occurs on their planet. First, participants completed a Baseline trial in which *shared* vs. *distinct* social group

membership was the only manipulated relational property. The latter two test trials consisted of an Aligned trial and a Crossed trial, where the manipulation of *close* vs. *distant* proximity properties was added (i.e., whether they all attend the same school, or each attend a different school, respectively). See Figure 6.

Figure 6. *Consensus Group Relational Properties*



Note. Relational properties manipulated across two social dimensions. Group membership dimension on the y-axis, ranging from *shared* to *distinct* group membership relations between consensus members. Proximity dimension on the x-axis, ranging from *close* to *distant* proximity relations between consensus members.

While a forced-choice measure was used in the preliminary studies, such a measure would not allow for the degree of relative preference for one consensus over another to be captured. Thus, the key dependent measure used across all trials was on a 6-point scale, enabling participants to indicate their level of confidence in the group that they chose as being the better

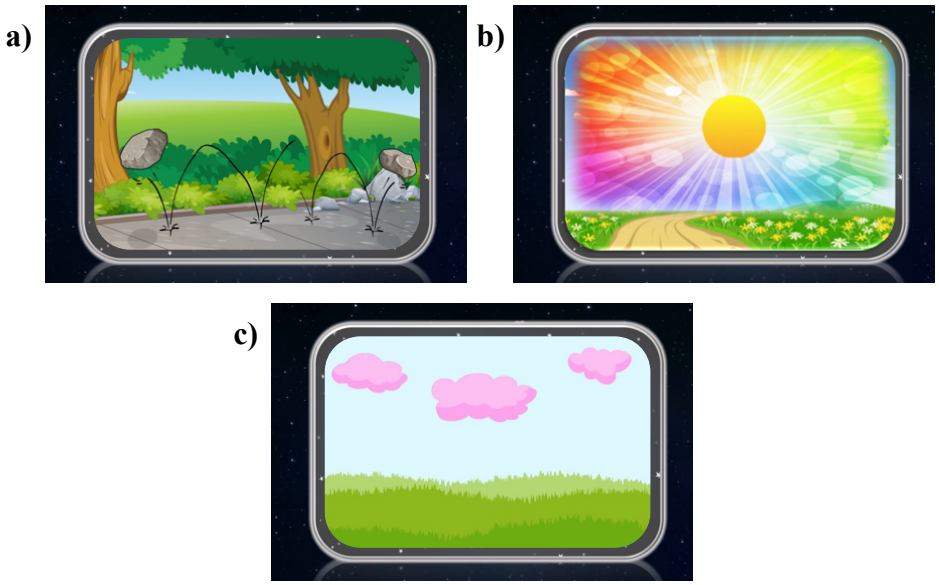
group to ask (i.e., “might be better,” “probably better,” “definitely better”). Additionally, participants were asked to explain their reasoning after reporting their relative preferences, which provided insight into the explanatory process underlying their choices.

Materials

Stimuli consisted of a set of PowerPoint slides which illustrated the story participants heard. For child participants, all sessions were conducted online over Zoom with a trained experimenter. Whereas for adults, sessions were conducted asynchronously over the Prolific online research platform, with stimuli presented via video recordings of narrations and accompanying visual animations.

The same alien stimuli from the preliminary studies were used. Three stimulus test items were designed, each referencing a natural phenomenon or “weird” thing that happens on Planet Kazar (See Figure 7). The three test items were written to have equal word counts and designed to refer to natural phenomena that is directly observable, and thus could be clearly depicted by a visual illustration. Additionally, four sets of three visuals for the (*close* vs. *distant*) proximity variable were created. Of these, two sets of visuals were designed to illustrate the “close” proximity property; each comprised of three identical illustrations of the same school and its surrounding natural landscape (See Figure 6 for example stimuli above). The remaining two sets were designed to illustrate the “distant” proximity property; each comprised of three illustrations of three different schools with varying natural landscapes (e.g., coastal, desert, mountain landscape).

Figure 7. *Test Trial Items*



Note. Stimulus test items used in the experiment, each referencing a natural phenomenon or "weird" thing that happens on Planet Kazar. a) "Look! Everywhere on Planet Kazar, the rocks are bouncy!"; (b) "Look! Everywhere on Planet Kazar, the sun shines rainbows!"; and (c) "Look! Everywhere on Planet Kazar, the clouds are pink!"

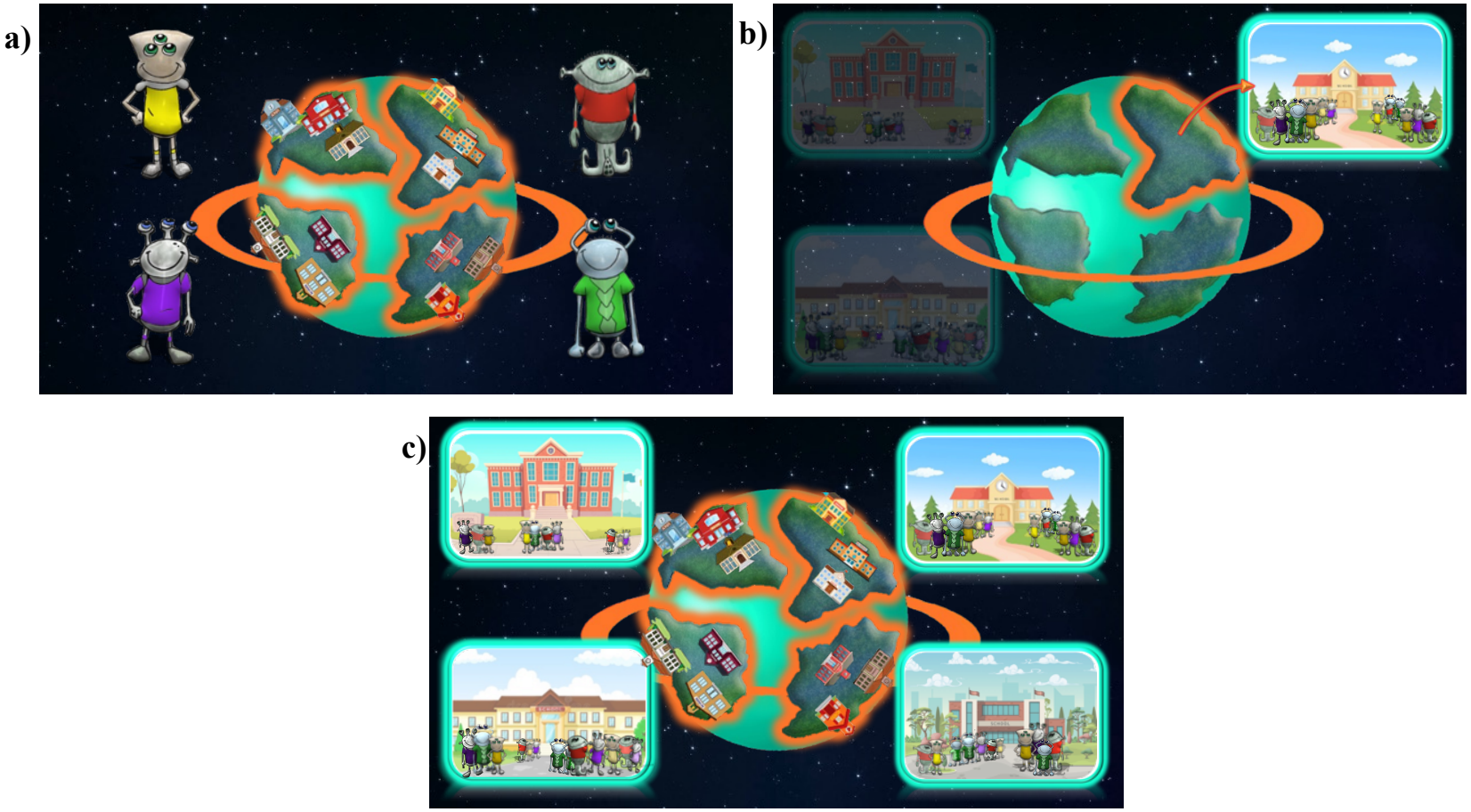
Two confidence scales were designed, one for children and the other for adults. Both scales included circles with relative sizes corresponding to each confidence level option, with larger circles aligning to greater confidence. The child scale included gender-matched illustrations to provide visual aids for comprehension with one depicting the highest confidence pole and the other marking the lowest confidence pole (See Figure 14 in Data Coding & Reliability section).

Procedure

Story Introduction

The experimenter began by introducing participants to a faraway planet named Planet Kazar. Participants were told that there are four kinds of aliens that live on Planet Kazar. The experimenter introduced each alien-kind using a generic noun label (e.g., Flurps) as an individual alien member of the respective alien group appeared on the screen (See Figure 8a). Next, participants were told that "*everywhere on Planet Kazar, you can find Flurps, Zarpies, Blarks, and Gorps going to school, just like you do!*" An orange border highlighted each continent on Planet Kazar while a visual depicting alien members of all kinds surrounding a school building appeared on the screen. The experimenter narrated, "*Look! See some Flurps, Zarpies, Blarks, and Gorps are going to school over here, and over here*. *And look they are going to school over here too*, *and over here!*" (See Figure 8b). Then, the experimenter repeated that, "*everywhere on Planet Kazar, you can find Flurps, Zarpies, Blarks, and Gorps going to school, just like you do!*" (See Figure 8c).

Figure 8. *Story Introduction Stimuli*

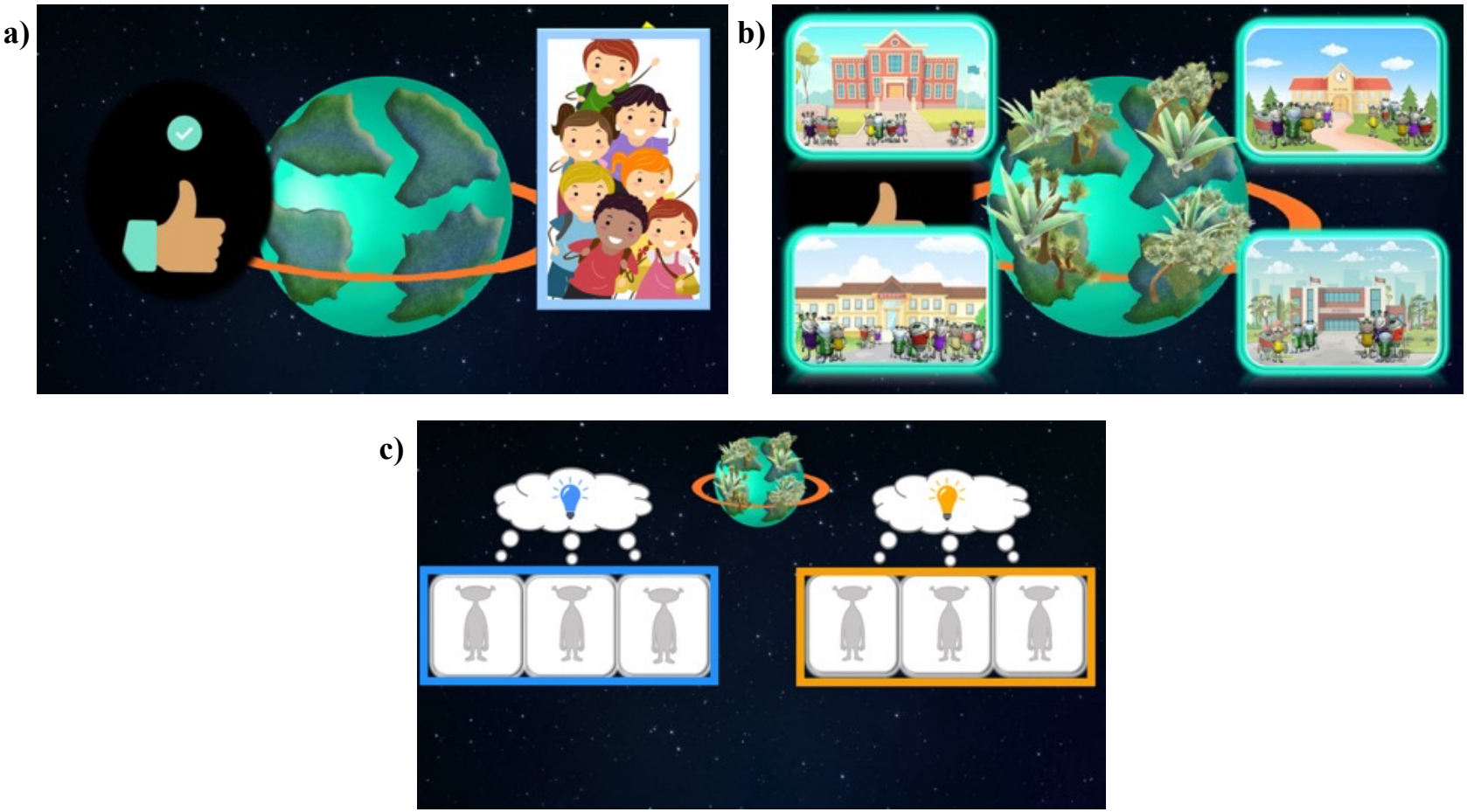


Note. The story introduction. a) Visual of individual alien members of the four cultural groups; b) School scene visuals to illustrate aliens attending schools on each of the four continents on Planet; and c) All school scene visuals.

Task instructions. The experimenter explained that because no one on Earth has ever seen or heard about Planet Kazar, they must learn the *right* things so that they would be able to teach everyone on Earth all about it (See Figure 9a). Participants were told that "*there are 'weird' things that happen in nature, everywhere, on Planet Kazar and so* [their] *job is to find out why those weird things happen*," and that the aliens living on the planet were there to help them do so (See Figure 9b). The four school-scene illustrations previously shown appeared once more to accompany the narration.

The experimenter proceeded to explain, "*First, I will tell you about a 'weird' thing that happens in nature, everywhere, on Planet Kazar. And then, I will show you some aliens*." Two groups of three alien-silhouettes appeared, one on the left and another on the right side of the screen. Following, the experimenter continued, "*now, what these aliens think* [left] *and what these aliens think* [right] *about why that weird thing happens is different*!" Above each respective group, a thought bubble appeared with a blue- and orange-colored light bulb, respectively, inside (see Figure 9c). Continuing the narration, the experimenter said, "*so to make sure you learn the right things, you need to figure out which group is the better group to ask why that weird thing happens*."

Figure 9. *Task Instructions Stimuli*



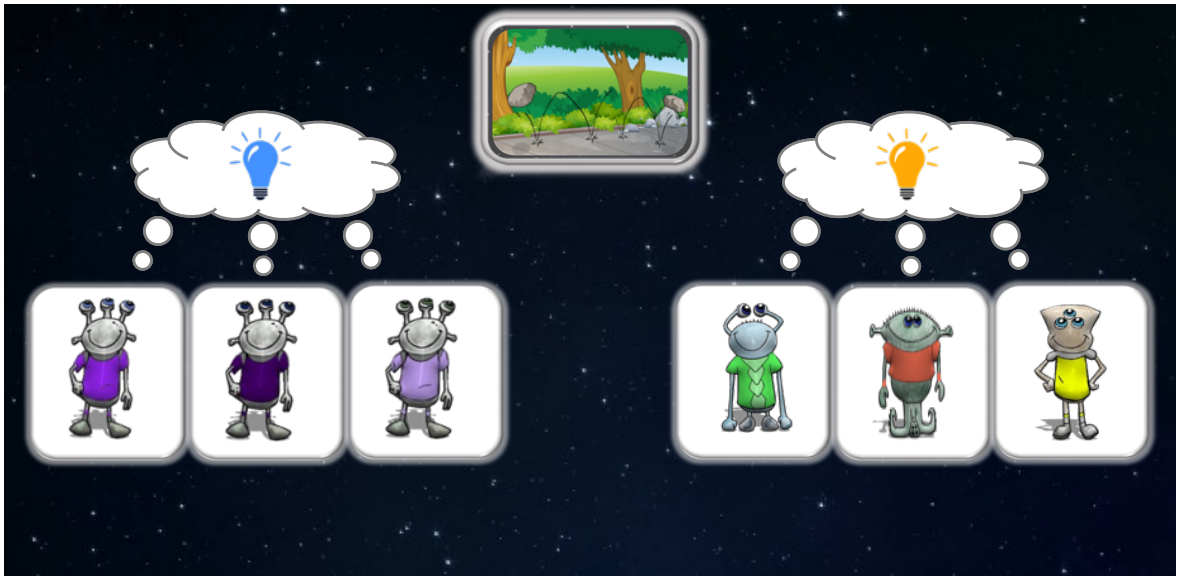
Note. Task instructions stimuli accompanying animations. a) Task instructions to learn the right things in order to help teach everyone on Earth; b) There are "weird" things that happen in nature everywhere on Planet Kazar, and the aliens are there to help [participants] learn the right things; c) Two groups of alien-silhouettes, each of whom think something different.

Test Phase

Participants completed three test trials: 1) Baseline trial; 2) Aligned trial; and 3) Crossed trial. On each trial, participants reported their relative preference for asking one of two consensus groups: one consensus with *shared* group membership (e.g., three Zarpies) and one with *distinct* membership (e.g., one Blark, one Gorp, one Flurp). The key difference across trials were the social compositions of the consensus groups between which participants chose.

Baseline Trial. The social compositions of the two consensus groups was identical to that of the preliminary studies. Specifically, participants chose between one consensus with *shared* group membership and one consensus with *distinct* membership See Figure 10 for example stimuli.

Figure 10. *Example Consensus Group Stimuli for the Baseline Trial.*



Note. Right: Consensus group with *shared* social group membership relational property. Left: Consensus group with *distinct* social group membership relational property.

Main Test Trials. Following the Baseline trial, the (*close* vs. *distant*) proximity property was added for the latter two test trials.

Aligned Trial. For the Aligned trial, the *shared* and *distinct* group membership property corresponded to the *close* and *distant* proximity property, respectively (See Figure 11).

Figure 11. *Example Consensus Group Stimuli for the Aligned Trial.*

Note. Right: Consensus group with *shared* social group membership and *close* proximity. Left: Consensus group with *distinct* social group membership and *distant* proximity.

Crossed Trial. For the Crossed trial, the consensus groups displayed the opposite combination of the two social relational properties: one consensus among individuals with *shared* group membership but with *distant* proximity, and one among individuals with *distinct* group membership but *close* proximity (See Figure 12).

Figure 12. *Example Consensus Group Stimuli for the Crossed Trial.*

Note. Right: Consensus group with *shared* social group membership and *distant* proximity relational properties. Left: Consensus group with *distinct* social group membership and *close* proximity relational properties.

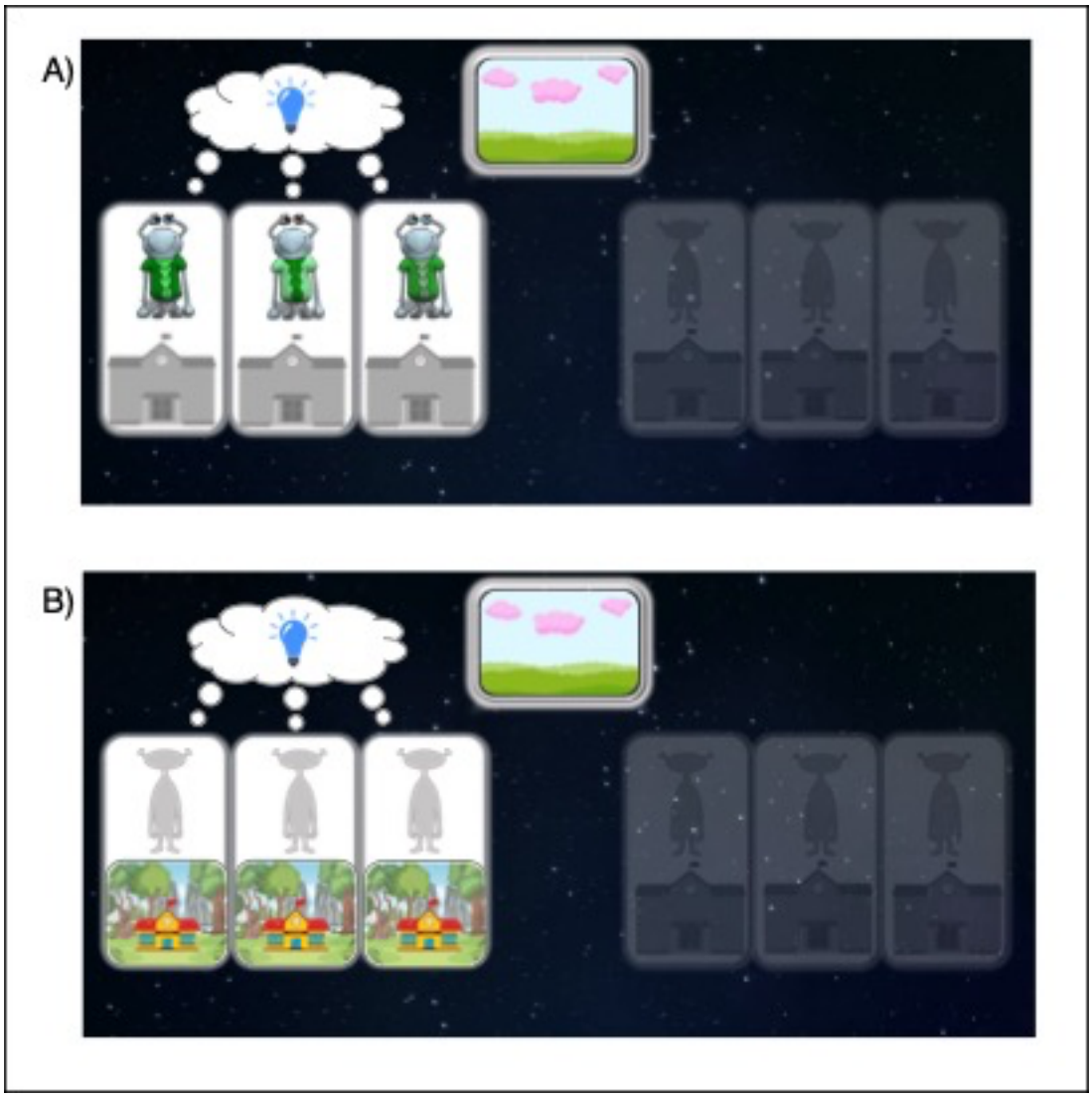
General Test Trial Procedure. The same general procedure was used for each trial. The experimenter began by stating, "*Now, I am going to tell you about the* [first/second/last] *weird thing that happens in nature everywhere on Planet Kazar."* A visual of the test item then appeared on the screen as the experimenter continued, "*Look! Everywhere on Planet Kazar,* [e.g., the sun shines rainbows]*!"* Next, the item-visual reduced in size and moved to the top-center of the screen to serve as a memory aid, and the experimenter continued, "*so, here are some aliens*," and two groups of three alien-silhouettes appeared, along with three school-silhouettes on the Aligned and Crossed trials.

The group on the right side of the screen then became transparent to direct participants attention to the group on the left side. The experimenter stated, "*Now <u>these</u> aliens all think the same thing about why the* [e.g., rocks are bouncy]," and a thought bubble appeared above the

group inside of which a blue [vs. orange] solid-colored light bulb to represent the consensus members' belief. Following, the experimenter revealed the group membership (*shared* vs. *distinct*) and proximity (*close* vs. *distant*) relational properties (for the Aligned and Crossed trials only). For counterbalancing purposes, the script was designed such that each relational property had a corresponding statement so the order in which the two relational properties were introduced could easily be manipulated while controlling for the specific language used.

The "group-statement" specified the group membership relational property; "*And look! This is a [Zarpie], this is a [Blark], and this is a [Gorp],*" or "*And look! This is a [Flurp], this is a [Flurp], and this is a [Flurp]*." As the "group-statement" was made, each alien member appeared on the screen and replaced one of the respective consensus group's alien-silhouettes (See Figure 13A). On the Aligned and Crossed trials, a "school-statement" specified the proximity relational property; "*And look! They each go to a different school*," or "*And look! They all go to the same school!*" As participants heard the "school-statement," three visuals appeared on the screen to replace the three school silhouettes (See Figure 13B). The two sets of three school environments comprised of three unique school contexts (*distant* proximity) and three identical illustrations of a single school context (*close* proximity).

Figure 13. *Test Trial Procedure: Group Membership & Proximity Relational Properties*



Note. Example stimuli: Introducing consensus group relational properties. (A) Example *shared* group membership relational property stimuli replacing respective consensus group's alien-silhouettes. (B) Example *close* proximity relational property stimuli replacing respective consensus group's school-silhouettes.

After completing the same procedure for the group on the right side of the screen, the experimenter then continued, "*Now remember, what these aliens think* [left] *and what these aliens think* [right] *about why* [e.g., the rocks are bouncy] *is different*!" To highlight the consensus groups' diverging views, the colors of the thought-bubble and light bulb above each respective group then inverted, such that the color of the thought-bubble changed from white to blue [orange] and the light bulb inside from blue [orange] to white. Finally, participants were asked the test question, "*So, which group is the better group to ask, why are* [e.g., the rocks are bouncy] *on Planet Kazar?*" As the experimenter continued to ask, "*Do you want to ask these aliens or these aliens?*" a solid-blue [-orange] rectangle appeared framing each respective consensus group.

Participants rated whether the consensus with the *shared* membership or the *distinct* membership relational property "might be better, is probably better, or definitely better to ask," producing a 6-point scale of relative preference, where 1 corresponds to "definitely the consensus with *shared* membership," and 6 corresponds to "definitely the consensus with *distinct* membership." Lastly, participants were asked to explain their answers after providing their rating on each trial. Children explained their answers verbally, whereas adults were asked to write two or three sentences to explain their reasoning.

Counterbalancing

Eight counterbalancing orders were created to vary the order of the natural phenomena test-items, as well as the order in which the latter two test trials were presented. The particular social group (i.e., Flurps, Zarpies, Blarks, or Gorps) and left/right presentation of the *shared/distinct* membership consensus groups presented on each trial were also counterbalanced between participants. Moreover, in order to control for potential recency effects of the two

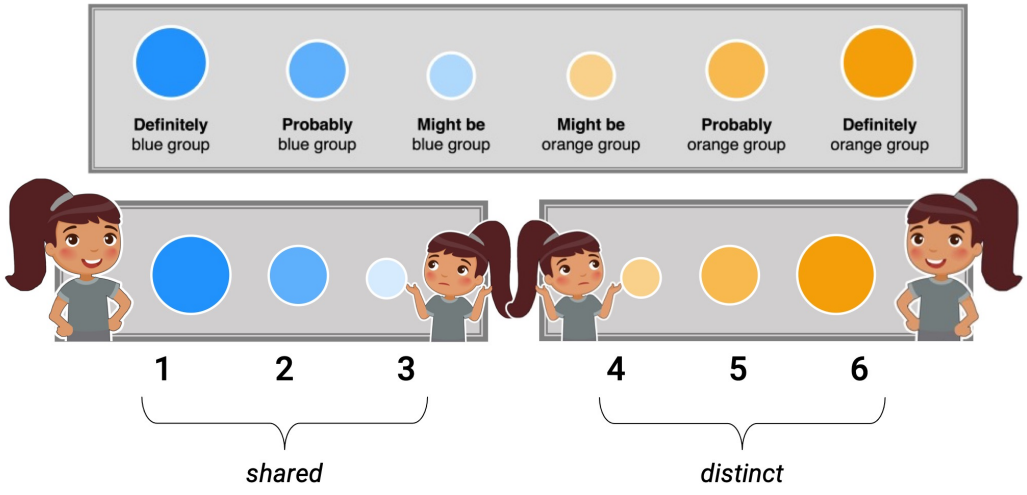
manipulated variables, the order of the "group-statement" and "school-statement" was counterbalanced between participants. Thus, two versions of each counterbalanced order were created, with a "Groups-First" version and a "Schools-First" version, resulting in a total of sixteen counterbalancing orders.

Data Coding & Reliability

Relative Preference Ratings

As specified above, on each trial, participants ratings were coded on a 6-point ordinal scale of relative preference from 1 = Definitely [2 = Probably; 3 = Might be] the consensus group with *shared* group membership to [4 = Might be; 5 = Probably] 6 = Definitely the consensus group with *distinct* group membership. Adults used the scale directly; whereas children's responses were staggered: they were first asked to make a forced choice decision between the *shared* vs. *distinct* membership consensus groups, and then were asked to make a "might be/probably/definitely" judgment (see Figure 14).

Figure 14. *Scale of Relative Preference*



Note. Rating scale of relative preference for shared and distinct group membership between consensus members. Top Row: Rating scale for adult participants. Bottom Row: Rating scale for child participants.

Reasoning

After providing their relative preference ratings on each trial, participants were asked to explain their reasoning. For example, after children responded to the question, "*So do you think that the orange group might be better to ask, is probably better ask, or is definitely the better group to ask?*" children were asked, "*Why do you think that the orange group is probably better to ask than the blue group?*" Children's responses were transcribed verbatim from video recordings of the testing session and subsequently checked for accuracy by trained research assistants. Adults provided written explanations, as they participated asynchronously on the Prolific platform.

Coding. There can be different reasoning underlying the same patterns of responses across relative consensus group preferences. Preferences for social diversity among consensus members can be based on associative grounds, or alternatively, a more *causal* understanding of social diversity as a mechanism for consensus accuracy. An associative account could simply be based on the fact that greater social diversity is more representative of the target population for which the knowledge in question is expected to be distributed. In this view, children might be making an inductive inference based on the relationship between the diverse relational properties of consensus groups and their expectation that knowledge about the natural world is widely distributed on Planet Kazar. However, a more sophisticated *causal* understanding of social diversity as a mechanism for consensus accuracy goes beyond inductive inference, to a consideration of *how* social diversity generates epistemic value.

The coding scheme shown in Table 1 was developed to highlight this distinction and examine developmental change, from a sense that more diverse relations offer better evidence for a claim to consideration of how social diversity generates epistemic value. Thus, participants'

explanations were classified into one of two overarching themes: 1) *Relations* (i.e., explanations that reference the relations between consensus members); and 2) *Mechanism* (i.e., explanations that reference the relations between consensus members as a mechanism for consensus accuracy). These two themes were the primary measure used for analysis.

Within each theme there are more granular conceptual categories. While these categories were not used for analyses, this granular coding was meant to provide descriptive information. Participants' explanations often met criteria for multiple coding categories. Thus, the final codes assigned to each explanation reflect the highest coding category for which the explanation met criteria. It is important to note that while the granular conceptual categories are hierarchically organized, no strong claims are being made regarding these distinctions within the themes. If participants' responses did not fit into one of the coding categories, their responses were coded as Ambiguous/Irrelevant. A separate coding category was used to indicate participants who reported "I don't know" or "I'm not sure" for a given trial.

Inter-rater reliability. Coding was completed by two coders. The first coder coded all explanations, and the second coder coded 20% of the participants' explanations across the three trial types. Percent agreement was used as a measure of reliability. For the more granular coding categories, there was 86.7% agreement for both the Baseline and Aligned trials, and 90% agreement on the Crossed trial. There was 100% agreement for the explanation coding for the main themes across the three trials. Disagreements in reference to the more granular coding categories were resolved by discussion.

Table 1. *Explanation Coding Scheme & Examples*

Theme	Category	Examples
Relations		
Reference to relations between…		
	1 Kinds.	• "Because they are different kinds of aliens."
	2 Perspectives. *thoughts/reasons/ideas.*	• "Because they're all different kinds, they might have all different kinds of ideas."
	3 Experiences. *places/knowledge-access.*	• "They all go to different schools, so they all see different things."
	4 Representation. *group/place exemplars.*	• "I think this group is better because if we were able to split one hundred into four, because there's four types, they would get 75% and this group would get 25%. Only the [Barks] think blue but this group is more because I get more [aliens] to ask."
Mechanism		
Reference to relations as mechanism to…		
	5 Pooled knowledge. *aggregating distinct perspectives/knowledge.*	• "Because this group has 3 different species, so it has more input into what the answer is, like more ideas of what it could be." • "Because they go to different schools so learn different things…then they can put it together and figure out the real reason." • "Because they came from different schools and that probably means the schools teach different things, and they have different knowledge and then they put that together to make more knowledge. If you have a lot of knowledge, it's better than just a little. If you know about a lot of things like science and math, it's better than just knowing about reading."
	6 Common knowledge. *spread of normative standards of belief.*	• "Because it seems like a lot of aliens think that because they each go to different schools and they're each different aliens." • "Because they go to different schools and their teachers would teach them that and it's three different teachers versus one teacher."
	7 Consilience. *independence between converging sources.*	• "Because they also go to different schools and have different ideas and are taught differently, so if they agree on something that probably means it's generally correct." • "Because this group has multiple kinds of aliens so they might have a different way to think of it…all ends up the same way, and the other ones all think the same because they're the same type of alien." • "They're different aliens [and] they're probably taught different from their parents. They go to very different schools, so they're obviously taught very different…If you have different opinions that means that you can form a better hypothesis than if you have all the same opinion… If you had all the same knowledge, you're basically just repeating … like there's nothing new about it and it might not be right. Like, for example if someday… you move to a new place and they taught you that the earth is flat…, but you were already taught that the earth is round. So, if they all go to the same school, it has a not a better chance of being correct, [but] more chance of it being incorrect."

Data Analytic Plan

The following sections will detail the data analytic models that were used to analyze participants' relative preference ratings and explanations in the present dissertation.

Preliminary Analyses

Preliminary analyses were conducted to examine if there were any differences in participants' response patterns across the eight counterbalanced orders, as well as across order-type (i.e., Groups-First vs. Schools-First) in the latter two test trials. However, because participants were randomly assigned to order-types and counterbalanced orders and assignment is balanced within each age group, significant differences between presentation-orders were not expected.

Primary Analyses.

General Analytic Method for Aims 1 & 2. The response variable for the analytic models used to address Aims 1 & 2 was a 1-6 ordinal rating scale (participants' preference ratings; ranging from 'Definitely the consensus with *shared* membership' to 'Definitely the consensus with *distinct* membership). Using standard regression models which treat ordinal scales as metric data has historically been a common approach to analyzing such data in psychology research. However, recent publications strongly advise against treating ordinal scales as metric data, demonstrating that treatment of ordinal data as metric increases risk of both Type I and Type II errors (Liddell & Kruschke, 2018) compared to alternative analyses. Thus, cumulative ordinal regressions with logit link functions were used to address Aims 1 and 2.

Rather than assuming that the distance between each level of response to an ordinal scale is equal, ordinal models assume that responses are derived from an ordered set of scale

categories (here being the relative preference ratings), which are mapped from an underlying continuous distribution, and search for the latent thresholds that "cut" the continuous distribution into response categories. An intuitive way to think of these models is to think of an ordinal regression as a set of logistic regressions, moving from one cut threshold to the next in order. Thus, in addition to producing the odds ratios for each coefficient in the model, an ordinal regression produces *k-1* thresholds, where *k* is the number of categories in the ordinal (e.g., Likert) scale. It's important to note that while the standard coefficients reflect change across the ordered set of scale categories (here being the relative preference ratings), the threshold cuts are the same for all levels of all variables included in the model. Thus, the threshold cut between levels 3 and 4 of the 6-point Likert scale of relative preference ratings can be exponentiated to compare ratings to chance. The ordinal package in R 4.2.3 (R Core Team, 2023), was used to the analyses specified below for each aim.

Aim 1 Hypotheses. Examine children's sensitivity to diversity based solely on social group membership. The first aim was to test whether age-related trends in children's tendency to prefer a diverse consensus group, suggested by the results of the preliminary studies discussed in Chapter 2, holds when group membership has no bearing on the proximity between consensus members. It was expected that around age 8 children will prefer the diverse consensus group (i.e., consensus with *distinct* group membership) more often than expected by chance, and with relative confidence ratings increasing with age.

Aim 1 Analysis. Cumulative ordinal regressions with logit link functions were performed with participants' relative preference ratings as the dependent variable and age-in-months (mean-centered) entered as the predictor (Model 1). Follow-up analyses were then conducted within

each age group, exponentiating the threshold cut between levels 3 and 4 of the 6-point Likert scale of relative preference ratings to compare ratings to chance (Model 2).

(Model 1): `clm(as.ordered(Rating)~Months_c.,data=.,link="logit")`

(Model 2): `clm(as.ordered(Rating)~1,data=.,link="logit")`

Aim 2 Hypotheses. Examine developmental change in children's preference for social diversity between converging sources. It was hypothesized that if children's judgments of reliability based on group membership reflect expectations of overlap (or lack thereof) in consensus members' individual experiences, then the following patterns of preference ratings across trials would be observed: 1) on the Aligned trial, it was expected that children's tendency to prefer a consensus with *distinct* group membership would remain consistent with ratings on the Baseline trial, as the group membership and proximity relational properties are in alignment; and 2) on the Crossed trial, relative preference ratings for *distinct* group membership and *distant* proximity relations between consensus members on the Baseline and Aligned trials would shift in the opposite direction. This shift in preferring a consensus with *shared* group membership and *distant* proximity was expected given the more concrete implications of the proximity relational property for consensus members' experiences.

Aim 2 Analysis. The relationship between relative preference ratings and trial type, and whether the direction and strength of this relationship varies with age was analyzed using Model 3, with age-in-months (mean-centered), trial type (centered on the Baseline trial), as well as their interaction entered as fixed effects. Follow up analyses were also conducted within each age group (Model 4) to obtain the predicted probabilities that children would give a rating of 4 or

more (“Might be the consensus group with *distinct* group membership”) on the 6-point scale on each trial type.

(Model 3): `clm(as.ordered(Rating)~TrialType_c*Months_c.,data=.,link="logit")`

(Model 4): `clm(as.ordered(Rating)~TrialType,data=.,link="logit")`

Aim 3 Hypotheses. Examine the explanatory process underlying children’s consensus group preferences. It was expected that the proportion of explanations that fall under the “Mechanism” coding theme would increase with age. However, independent of age, it was expected for the tendency to provide explanations within the “Mechanism” coding theme to correspond to the extent to which participants’ patterns of ratings across trials conform to the developmental progressions outlined in Aim 2.

Aim 3 Analysis. The key dependent measure used to address Aim 3 was whether participants’ explanations fell under the “Relations” or the “Mechanism” coding theme. Difference scores in individual participants’ relative preference ratings between each trial type (Baseline vs. Aligned; Baseline vs. Crossed; Baseline vs. Aligned) were computed. Generalized Linear Mixed Models (GLMM) were used with explanation themes (Relation = 0; Mechanism = 1) entered as the outcome variable. Model 5 was used to analyze the relationship between explanation themes and the difference scores for each trial type comparison. The difference score for the trial type comparison, age-in-months (mean-centered), and their interaction were entered as predictors, and with random intercepts for participants. for each trial type comparison within each age group (Model 6).

(Model 5):
```
glmer(Theme ~ Diff_Score*Months_c.+(1|Subject),data=.,
                        family = binomial('logit'))
```

Chapter 4: Results

The results of the present study are conceptually organized into subsections corresponding to each of the primary aims of this dissertation. Per the data analytic plan outlined in the previous chapter, the first section of results will include preliminary analyses conducted. The next section of results will address Aim 1: Examine children's sensitivity to diversity based solely on social group membership. The following section will address Aim 2: Examine developmental change in children's preference for social diversity between converging sources. The final section will address Aim 3: Examine the explanatory process underlying children's consensus group preferences.

Preliminary Analyses

Preliminary analyses were conducted to examine if there were any differences across order-type (i.e., Groups-First vs. Schools-First) in participants' ratings across the latter two test trials. A cumulative ordinal regression was fit with the full sample, and only the child sample; both models results indicated that there were no differences by order-type in participants' relative preference ratings on the Crossed and Aligned trial types (Log-$OR_{child+adult}$= -0.03, *SE* = 0.18, *z* = -.18, *p* =.86; Log-$OR_{child.}$= -.14, *SE* = 0.21, *z* = -.67, *p* = .50). Therefore, order-type was not included as a covariate in subsequent analyses.

Primary Analyses

Aim 1. Examine children's sensitivity to diversity based solely on social group membership.

To test whether age-related trends in children's tendency to prefer a diverse consensus group holds when group membership has no bearing on the proximity between consensus members, a cumulative ordinal regression with logit link function was performed with participants' relative preference ratings as the dependent variable and age-in-months (mean-centered) entered as the predictor.

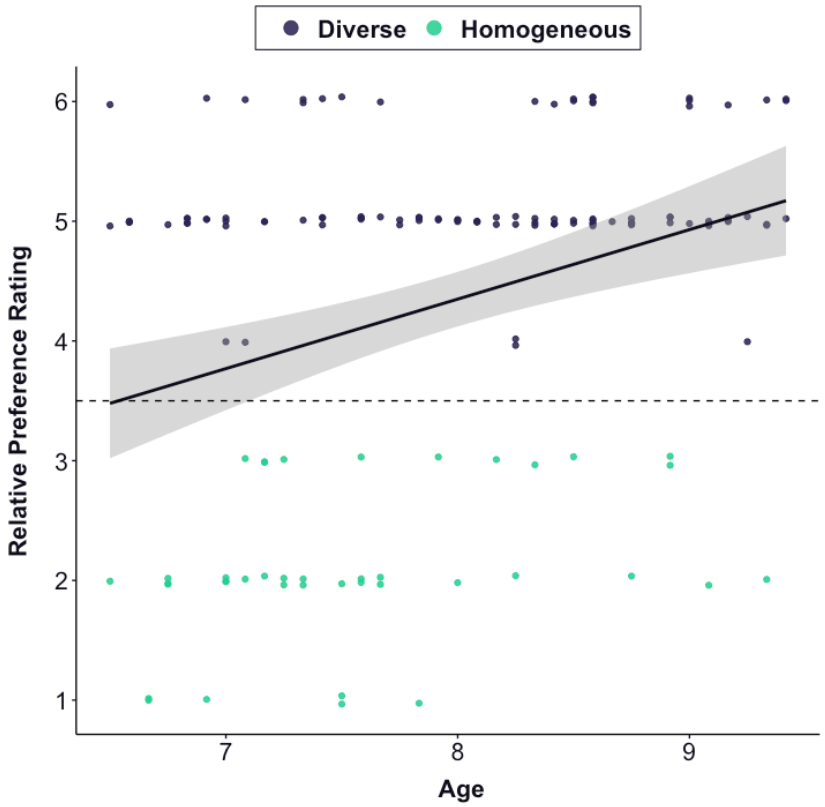
Results from a likelihood ratio test suggest that including age-in-months significantly improved model fit, as compared to the null model, $\chi^2_{(1)}$= 15.05, $p < .001$. There was a main effect of age, ($OR_{Age.c}$= 1.07, 95% CI: [1.03, 1.10], $p < .001$), suggesting children's relative confidence that the diverse consensus group (i.e., consensus with *distinct* group membership) is the *better* group to ask increased by 10.7% with every month increase in age.

As outlined in the data analytic plan section in the previous chapter, while the standard coefficients reflect change across the ordered set of scale categories (here being the relative preference ratings), the threshold cuts are the same for all levels of all variables included in the model. Thus, the threshold cut between levels 3 and 4 of the 6-point Likert scale of relative preference ratings can be exponentiated to compare ratings to chance. Secondary analyses within each age group were performed to compare ratings to chance.

Results suggested that on the Baseline trial, 8- and 9-year-olds, but not the 7-year-olds, rated the diverse consensus group as the *better* group to ask more often than expected by chance. Specifically, the model predicts that 41.4% of 7-year-olds (z = -1.22, 95% CI:[28.8, 55.2]), 28.8% of 8-year-olds (z = -4.67, 95% CI:[21.6, 37.1]) and 12.2% of 9-year-olds (z = -4.46, 95% CI:[5.7, 25.1]), would report a relative preference rating of 3 or *less* ("Might

be/Probably/Definitely the consensus group with *shared* group membership") on the 6-point scale. Thus, children's tendency to prefer the diverse consensus group beginning around age 8 demonstrated by the preliminary studies' results was replicated when the confound between group membership and proximity (geographic location) was removed. This suggests that children's sensitivity to diversity based on group membership alone is sufficient as the basis for children's evaluations.

Figure 15. *Relative Preference Ratings on the Baseline Trial by Age*



Note. Relative preference ratings for a consensus with diverse or homogeneous group membership relational property on the Baseline trial, plotted as a function of age, along with a logistic regression fit to the dataset. Points are jittered along the Y axis (but not the X axis). The gray band shows a 95% confidence interval in the regression.

Aim 2. Examine developmental change in children's preference for social diversity between converging sources.

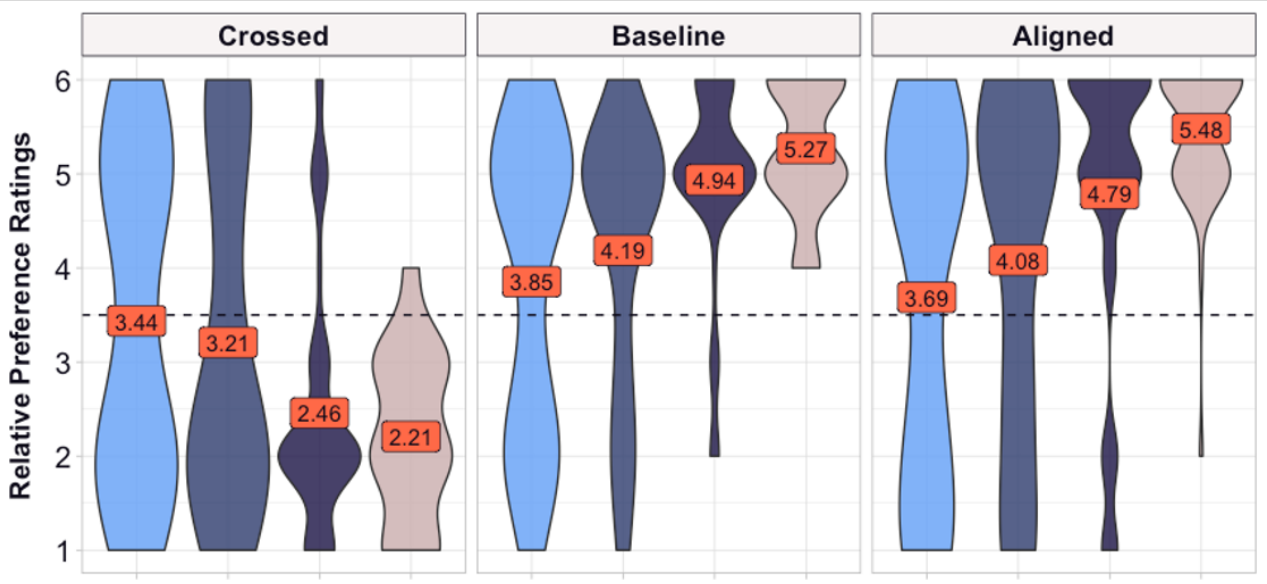
A cumulative ordinal regression with logit link function was performed with age-in-months (mean-centered), trial type (centered on the Baseline trial), as well as their interaction entered as predictors. Results from likelihood ratio tests suggest that the predictors significantly improved model fit, as compared to the null model, $\chi^2_{(3)}$=74.14, $p < .001$. Figure 16 shows the distribution of children's preference ratings across trials.

There was a main effect for age ($OR_{Age.c}$= .96, *SE* = .01, *z* = -2.80, 95% CI:[.93, .98], $p <$ .001). Across all age groups, participants' relative preference for the consensus with *distinct* group membership on the Crossed trial increased by a factor of 3.34 on the Baseline trial ($OR_{Baseline}$= 3.34, *SE* = .22, *z* = 5.61, 95% CI:[2.19, 5.08], $p < .001$), with the difference between levels of reported ratings on the Baseline and Cross trials increasing with age ($OR_{Age.c*Baseline}$= 1.11, *SE* = .02, *z* = 4.63, 95% CI:[1.06, 1.15], $p < .001$). Thus, children overall were more likely to prefer the consensus with *distinct* group membership on the Baseline trial, than on the Crossed trial, and this tendency increased with age.

Likewise, across all age groups, participants' relative preference for the consensus with *distinct* membership on the Crossed trial increase by a factor of 3.54 on the Aligned trial ($OR_{Aligned}$= 3.54, *SE* = .23, *z* = 5.60, 95% CI:[2.27, 5.51], $p < .001$). Moreover, the magnitude of this difference between ratings on the Crossed and Aligned trials also increased with age ($OR_{Age.c*Aligned}$= 1.09, *SE* = .02, *z* = 4.23, 95% CI:[1.05, 1.14], $p < .001$). Thus, again children overall were more likely to prefer the consensus with *distinct* group membership on the Aligned trial, than on the Crossed trial, and this tendency increased with age.

Thus, children's patterns of preference ratings across trials are consistent with the perspective that children's judgments of reliability based on group membership reflect expectations of overlap (or lack thereof) in consensus members' individual experiences. Children's tendency to prefer a consensus with *distinct* group membership on the Baseline trial remained consistent with ratings on the Aligned trial, when the group membership and proximity relational properties were in alignment. On the Crossed trial, with age, children's relative preference ratings shifted in the opposite direction, suggesting an increasing sensitivity to the more concrete implications of the proximity relational property for consensus members' experiences.

Figure 16. *Relative Preference Ratings by Trial Type*



Note. Violin plots and means for each trial type across age groups. Participants reported ratings of relative preference on a 6-point scale (1=Definitely the consensus with *shared* group membership [2=Probably; 3=Might be], to [4=Might be; 5=Probably] 6=Definitely the consensus with *distinct* group membership).

Follow up analyses within each age group were conducted to obtain the predicted probabilities that children would give a rating of 4 or *more* ("Might be the consensus group with *distinct* group membership") on the 6-point scale on each trial type. Table 2 shows the predicted probabilities obtained for each trial type across age groups. Consistent with the pattern of results reported above, with age, children's tendency to preference for *distinct* group membership among consensus members increased on the Baseline and Aligned trials. Moreover, with age, children's relative preference for the consensus with *distinct* group membership decreased on the Crossed when there was *close* proximity (i.e., attend the same school) between its members.

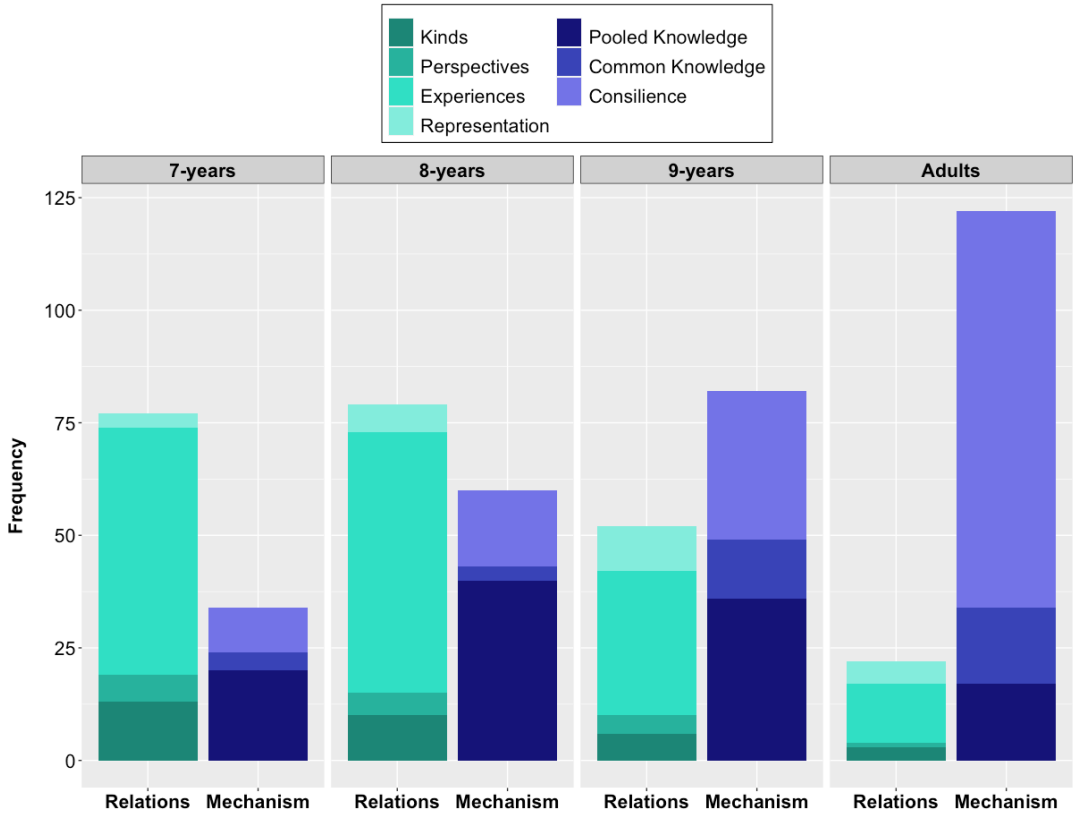
Table 2. *Predicted Probabilities of Consensus Group with Distinct Membership Preferences*

Age Group	Baseline trial		Aligned trial		Crossed trial	
	Predicted	95% CI	Predicted	95% CI	Predicted	95% CI
7-years	58.6%	[46.1, 71.1]	54.8%	[41.2, 68.4]	48.9%	[35.8, 62.0]
8-years	62.4%	[50.6, 74.3]	65.0%	[52.1, 77.8]	42.4%	[29.1, 55.8]
9-years	82.5%	[73.6, 91.4]	85.6%	[77.1, 94.1]	22.5%	[11.8, 33.2]
Adults	99.0%	[96.9, 100]	99.5%	[98.4, 100]	7.8%	[0, .15]

Aim 3. Examine the explanatory process underlying children's consensus group preferences.

Descriptive information. Of the 432 explanations provided by child participants across the three age groups, and three trial types, 25 explanations were coded as ambiguous or irrelevant (17.36%) and 12 explanations were coded under the "I don't know" category (2.7%). Figure 17 shows the distribution of the more granular codes assigned to participants' explanations within the "Relations" and "Mechanism" themes across age groups.

Figure 17. *Explanation Themes by Descriptive Codes*



Note. Proportions of descriptive codes within the Relations theme (green) and Mechanism theme (purple).

Trial Type Comparisons. Generalized Linear Mixed Models were performed for each trial type comparison, including random intercepts for participants. Explanation themes (Relation = 0; Mechanism = 1) were entered as the outcome variable, with the difference scores, age-in-months (mean-centered), and their interaction entered as predictors. Table 3 includes descriptive statistics for the difference scores computed for each trial type comparison.

Table 3. *Descriptive Statistics for Difference Scores Across Trial Type Comparisons*

Trial Type Comparison	*Mean*	*SD*	*Range*
Aligned-Baseline			
7-years	-0.17	2.23	[-5, 4]
8-years	-0.10	1.85	[-5, 4]
9-years	-0.15	1.51	[-5, 3]
Adults	0.21	.92	[-4, 2]
Baseline-Crossed			
7-years	0.42	2.57	[-5, 5]
8-years	0.98	2.93	[-5, 5]
9-years	2.48	1.96	[-3, 5]
Adults	3.06	1.33	[1, 5]
Aligned-Crossed			
7-years	0.25	3.20	[-5, 5]
8-years	0.88	3.44	[-5, 5]
9-years	2.33	2.38	[-4, 5]
Adults	3.27	1.22	[0, 5]

Aligned vs. Baseline trial comparison. Likelihood ratio tests suggest that including age significantly improved model fit, as compared to a null model $\chi^2_{(1)}$= 20.27, p < .001. Including the difference scores in the model did not improve model fit, $\chi^2_{(2)}$= 1.87, p = .39. The conditional R^2 for the full model (Model 2) indicated that 28.0% of the variance in children's explanations coded under each theme (Relations vs. Mechanism) was explained by both fixed and random effects included in the model. Table 4 includes results obtained for each model.

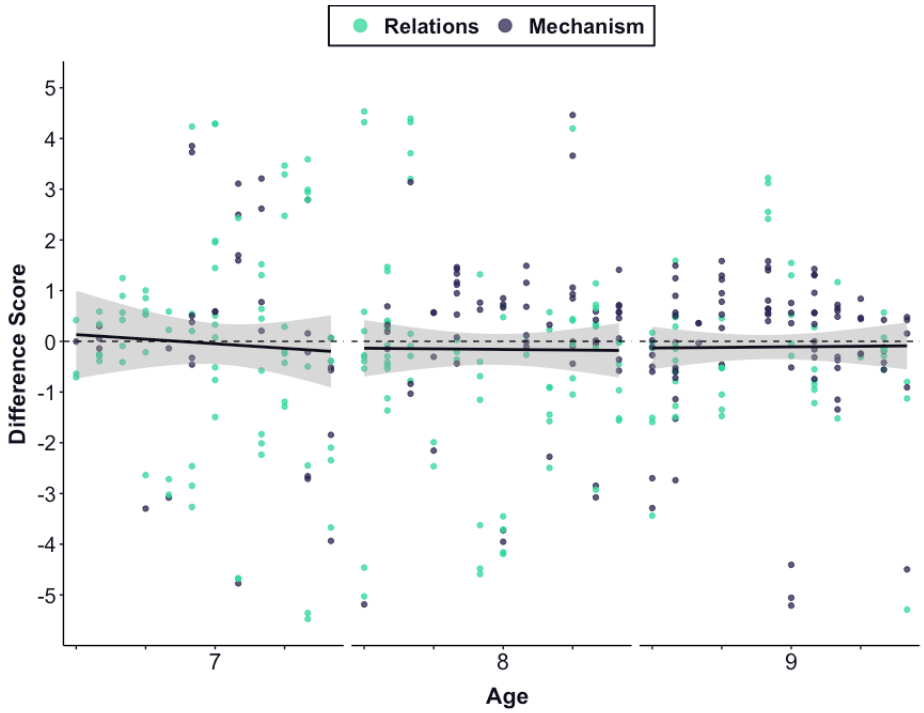
The difference scores variable was not significant (OR_{Diff}= .90, *SE* = .09, *z* = 1.21, *p* =.23), nor was the interaction between age and difference scores ($OR_{Diff*Age.c}$= .99, *SE* = .01, *z* = 0.77, *p* =.44). There was a main effect for age both when included as the single predictor in the model (Model 1: $OR_{Age.c}$= 1.08, *SE* = .01, *z* = 26.51, *p* < .001) as well as when the difference scores were added in the full model (Model 2: $OR_{Age.c}$= 1.03, *SE* = .02, *z* = 4.004, *p* < .001). These results suggest that irrespective of difference scores (i.e., difference in relative preference ratings on Aligned and Baseline trials), the odds of children's explanations falling under the mechanism coding theme increased by 10.8% with every month increase in age.

Table 4. *Aligned vs. Baseline Trial Comparison Analysis*

	Null Model		Model 1		Model 2	
Predictors	*OR*	*95% CI*	*OR*	*95% CI*	*OR*	*95% CI*
Intercept	0.84	0.61–1.16	0.78***	0.78–0.78	0.79	0.57–1.09
Age.c			1.08***	1.07–1.08	1.08***	1.04–1.12
Difference_Score.c					0.90	0.75–1.07
Age.c*Difference_Score.c					0.99	0.97–1.01
Random Effects						
τ_{00}	0.97 Subject		0.66 Subject		0.70 Subject	
ICC	0.23		0.17		0.18	
Marginal R^2 / Conditional R^2	0.000 / 0.227		0.117 / 0.265		0.127 / 0.280	

*$p<0.05$ **$p<0.01$ ***$p<0.001$

Figure 18. *Aligned vs. Baseline Difference Scores by Age & Explanation Theme*



Note. Aligned-Baseline difference scores plotted as a function of age, along with a logistic regression fit to the dataset. Points are color coded according to Explanation Theme and jittered along the Y axis (but not the X axis). The gray band shows a 95% confidence interval in the regression.

Baseline vs. Crossed trial comparison. Likelihood ratio tests suggest that including age significantly improved model fit, as compared to a null model $\chi^2_{(1)}$= 12.34, p < .001. However, including the difference scores in the model significantly improved model fit, as compared to the model with only age entered as the predictor $\chi^2_{(2)}$= 43.43, p < .001. The conditional R^2 for the full model (Model 2) indicated that 49.3% of the variance in children's explanations coded under each theme (Relations vs. Mechanism) was explained by both fixed and random effects included in the model. Table 5 includes results obtained for each model.

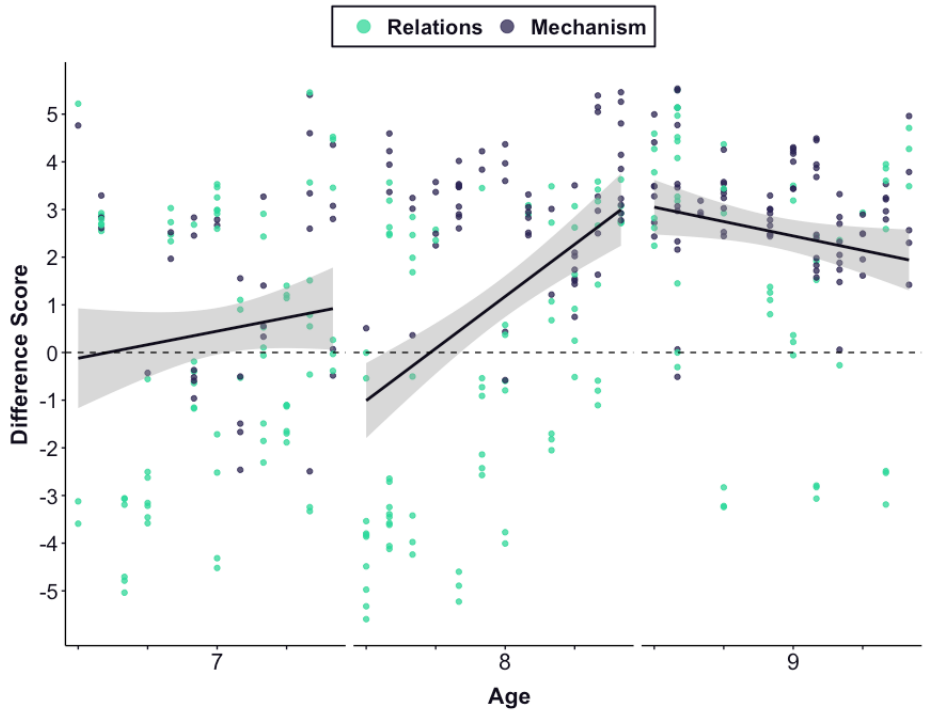
While there was a main effect for age when included as the single predictor in the model (Model 1: $OR_{Age.c}$= 1.07, *SE* = .02, z = 3.18, p < .01), it was no longer a significant predictor when the difference scores were added in the full model (Model 2: $OR_{Age.c}$= 1.03, *SE* = .02, z = 1.37, p = .17). There was a significant main effect of the difference scores (Model 2: OR_{Diff}= 1.67, *SE* = .11, z = 4.78, p < .001), suggesting that the odds that children provided explanations under the mechanism coding theme increased by a factor of 1.67 for every 1-point increase in difference scores of relative confidence ratings.

Table 5. *Baseline vs. Crossed Trial Comparison Analysis*

	Null Model		Model 1		Model 2	
Predictors	*OR*	*95% CI*	*OR*	*95% CI*	*OR*	*95% CI*
Intercept	0.71	0.48–1.05	0.67*	0.45–1.00	0.61*	0.41–0.92
Age.c			1.07**	1.03–1.12	1.03	0.99–1.07
Difference_Score.c					1.67***	1.35–2.06
Age.c*Difference_Score.c					0.99	0.98–1.01
Random Effects						
τ_{00}	1.95 $_{Subject}$		1.67 $_{Subject}$		0.89 $_{Subject}$	
ICC	0.37		0.34		0.21	
Marginal R^2 / Conditional R^2	0.000 / 0.373		0.084 / 0.392		0.356 / 0.493	

** p<0.05 ** p<0.01 *** p<0.001*

Figure 19. *Baseline vs. Crossed Difference Scores by Age & Explanation Theme*



Note. Baseline-Crossed difference scores plotted as a function of age, along with a logistic regression fit to the dataset. Points are color coded according to Explanation Theme and jittered along the Y axis (but not the X axis). The gray band shows a 95% confidence interval in the regression.

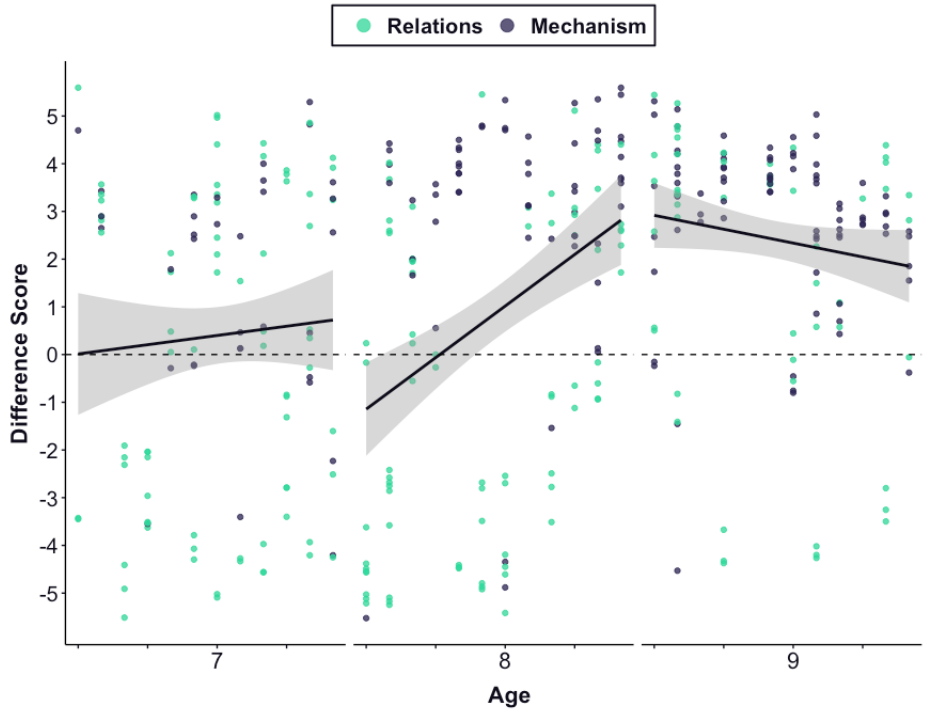
Aligned vs. Crossed trial comparison. Consistent with the pattern of results for the Baseline-Crossed trial comparison, likelihood ratio tests indicated that as compared to a null model, including age as a predictor significantly improved model fit, $\chi^2_{(1)}$= 19.32, p < .001. Likewise, including the difference scores in the model significantly improved model fit, as compared to the model with only age entered as the predictor $\chi^2_{(2)}$= 48.62, p < .001. The conditional R^2 for the full model (Model 2) indicated that 60.1% of the variance in children's explanations coded under each theme (Relations vs. Mechanism) was explained by both fixed and random effects included in the model. Table 6 shows the results obtained from each model.There was a main effect for age both when included as the single predictor in the model (Model 1: $OR_{Age.c}$= 1.11, *SE* = .03, z = 3.60, p < .001) as well as when the difference scores were added in the full model (Model 2: $OR_{Age.c}$= 1.05, *SE* = .02, z = 2.05, p = .04). There was a significant main effect of the difference scores (Model 2: OR_{Diff}= 1.69, *SE* = .11, z = 4.82, p < .001), suggesting that the odds that children provided explanations under the mechanism coding theme increased by a factor of 1.69 for every 1-point increase in difference scores of relative confidence ratings.

Table 6. *Aligned vs. Crossed Trial Comparison Analysis*

	Null Model		Model 1		Model 2	
Predictors	*OR*	*95% CI*	*OR*	*95% CI*	*OR*	*95% CI*
Intercept	0.75	0.47–1.19	0.71	0.44–1.14	0.57*	0.36–0.91
Age.c			1.11***	1.05–1.18	1.05*	1.00–1.10
Difference_Score.c					1.69***	1.37–2.09
Age.c*Difference_Score.c					1.01	0.99–1.03
Random Effects						
τ_{00}	3.83 $_{Subject}$		3.30 $_{Subject}$		1.65 $_{Subject}$	
ICC	0.54		0.50		0.33	
Marginal R^2 / Conditional R^2	0.000 / 0.538		0.140 / 0.571		0.401 / 0.601	

*$p<0.05$ **$p<0.01$ ***$p<0.001$

Figure 20. *Aligned vs. Crossed Difference Scores by Age & Explanation Theme*



Note. Aligned-Crossed difference scores plotted as a function of age, along with a logistic regression fit to the dataset. Points are color coded according to Explanation Theme and jittered along the Y axis (but not the X axis). The gray band shows a 95% confidence interval in the regression.

Chapter 5: Discussion

The current dissertation investigated the development of children's ability to reason about the link between individuals' experiences and the beliefs that they hold to evaluate reliability of converging sources. In contrast to prior work in which children were provided with *direct* information about overlap in consensus members' experiences, this dissertation study provides new insights on developmental trends in reasoning about *indirect* dependencies between consensus members based on sharing *common backgrounds*. As mentioned in Chapter 1, the central thesis guiding the current work is that building on the expertise of others inherently involves dependencies across sources. This makes it essential to investigate how indirect cues to the overlap of consensus members' social networks impact how children aggregate information from converging sources.

The interrelated social networks that comprise the human social world reflect overlapping channels of information, shaping the perspectives through which individuals interpret their experiences. From early on, children can reason about the implications of individuals' direct interactions with the world on the beliefs that they form (Aboody et al., 2022; Brosseau-Liard & Birch, 2011; Nurmsoo & Robinson, 2009; Robinson et al., 2011). Children and adults alike also use social group membership to make inferences about group members' experiences and, in turn, the beliefs they are likely to adopt (Karadağ & Sóley, 2022; Marble & Boseovski, 2019; Roberts et al. 2017; Sóley, 2019). Given that overlap of channels of information leads to redundancy, the ability to consider how social diversity generates epistemic value is a critical competency for learning about the world through the minds of others. One of the central aims of this dissertation was to disentangle the influence of social connections, indicated by shared group membership and close proximity, on children's judgements of reliability.

Children between the ages of 7 and 9, as well as adults, were introduced to a novel planet and alien groups that live there. Children were told that, much like children on earth, everywhere on the planet, aliens attend school. Tasked with learning the "right things" about why various natural phenomena occur on this planet, children were asked which one of two consensus groups, each of whom collectively thought something different, was the "better" group to ask. Across three trials, children chose between one consensus with *distinct* group membership and one consensus with *shared* group membership. Beyond social group membership, the key manipulation was *close* versus *distant* proximity between consensus members, with all consensus members attending the same school or each consensus member attending a different school, respectively. Children rated their relative preference for one consensus group over another using a 6-point scale and were asked to explain their reasoning.

Overall, these findings are consistent with the proposal outlined in Chapter 2, namely that qualitative changes in children's ability to consider how dependencies can lead to redundant information (thus reducing the evidential support for claims endorsed by converging sources) parallel the developmental shift in children's appreciation for interpretive diversity in middle childhood (i.e., iToM; See Carpendale & Chandler, 1996, 1998; Lalonde & Chandler, 2002; Ross et al., 2005). Specifically, the results from this study highlight that children between ages 7 and 9 increasingly recognize and flexibly utilize the relations between the proximity and social group membership of consensus members as a basis for evaluating reliability. The subsequent section outlines a summary of the findings across the three primary aims that guided this dissertation study.

Summary of Findings

On the Baseline trial, in which the consensus groups varied only on group membership, children's tendency to rate the consensus with *distinct* (as opposed to *shared*) group membership as the *better* group to ask increased with age. While 8- and 9-year-olds preferred the diverse consensus above levels expected by chance, 7-year-olds remained equally split in their preferences. Thus, consistent with the findings from the Preliminary Studies outlined in Chapter 2, by age 8 children can recognize the merits of diverse perspectives for consensus accuracy. On the latter two test trials, when the *close* and *distant* proximity relational properties were added, children showed an increasing sensitivity to social diversity based on both relational properties. Specifically, children's increasing tendency to prefer *distinct* group membership among consensus members on the Baseline trial, remained consistent on the Aligned trial when the *distinct* group membership and *distant* proximity relational properties were in alignment. On the Crossed trial, *close* proximity between members of a consensus with *distinct* group membership weakened children's relative preference for that group, with this tendency increasing with age.

Thus, by age 8, children reliably consider *distant* proximity and *distinct* group membership between consensus members as a cue to consensus reliability. In line with predictions, developmental trends in children's preference for *distant* proximity between consensus members remained consistent even when they shared group membership. While not directly examined in this study, children's explanations support the view that the proximity relational property was understood as reflecting more concrete evidence of (overlapping) experiences underlying agreement between consensus members.

Children were also asked to explain their choices on each trial. Their explanations were classified into one of two overarching themes, 1) *Relations* (i.e., explanations that reference the

relations between consensus members) and 2) *Mechanism* (i.e., explanations that reference the relations between consensus members as a mechanism for consensus accuracy). Given that different reasoning could underlie the same patterns of responses across relative consensus group preferences, this coding scheme was developed to examine qualitative change from a) a sense that the more diverse relations displayed by consensus members, the better evidence for a claim, to b) consideration of *how* social diversity generates epistemic value. The proportion of *Mechanism* explanations that children gave increased with age, independent of children's patterns of ratings. However, across all age groups, the degree to which children's ratings across trials reflected their weighing of the relative value of each relational property, correlated with an increasing tendency to provide *Mechanism* explanations for their choices.

Taken together, these findings demonstrate that by middle childhood, expectations for shared experience among members of social categories shape how children reason about the reliability of different patterns of consensus. Consistent with the perspective that children's evaluations reflect reasoning about the experiences that underlie agreement between consensus members, the developmental shift in children's preference for diversity reflects their ability to consider the epistemic value of social diversity *itself* for consensus accuracy. Given that we know it is also in middle childhood that children come to understand that differences in *perspectives* can lead to differences in beliefs (e.g., Chandler & Carpendale, 1998; Lalonde & Chandler, 2002; Ross et al., 2005), the current dissertation highlights an important link between understanding interpretive diversity and recognizing the epistemic value that social diversity generates. Nevertheless, another likely correlate is children's epistemological understanding—specifically, children's tendency to acknowledge that there can be multiple routes to knowledge,

as well as recognize some sources of knowledge as better than others (Heiphetz et al., 2013; Kuhn, 2020; Kuhn et al., 2000).

Epistemological Understanding as a Possible Developmental Mechanism

The developmental progression in children's ability to recognize the *relative evidential strength* of different patterns of consensus follows a similar trajectory to that of epistemological understanding proposed by Kuhn et al. (2000). On this view, the first level of epistemological understanding represents an *absolutist* conception of knowledge-- a belief that there is a single truth, and thus that all knowledge is knowable with certainty. Progressing further in development, at the *multiplist* level, knowledge is reconceptualized from an external objective entity to a product of the mind, open to interpretation. This leads to an understanding that it is possible for individuals to have different interpretations of the same event. Finally, progression to the *evaluativist* level involves increasing coordination of objective and subjective elements of knowing, coming to appreciate that two individuals can both have legitimate positions (both can be "right") but that one can have more merit (one can "be more right") based on the relative strength of the evidence supporting each position.

Children's increasing sensitivity to both the *distinct* group membership and *distant* proximity relational properties in middle childhood demonstrated by the current dissertation may reflect the developmental levels of epistemological understanding (e.g., Kuhn et al., 2015; Kuhn 2001). The pattern of ratings across trials consistent with an *absolutist* understanding reflects a preference for consensus members with the same experiences. The more perceived opportunities for overlap in experience, the more likely it is that agreement between consensus members reflects their shared access to that single truth. Thus, on the Baseline and Aligned trials, an *absolutist* understanding leads to a preference for *shared* group membership among consensus

members. Whereas, on the Crossed trial, this understanding leads to a shift in preference to a consensus with *distinct* group membership but *close* proximity, given the implications of shared experience among members of a consensus who all attend the same school. In the current study, of the 144 children in the sample, 22 children exhibited patterns of ratings across trials that reflect a systematic preference for shared experiences between consensus members. Of the 22 children, 19 justified their choices by referencing shared social relational properties and shared knowledge. The remaining three provided ambiguous or otherwise irrelevant explanations.

According to Kuhn, the hallmark of a *multiplist* understanding is a conception of knowledge as opinions of a subjective reality. It follows that in the context of the current dissertation study, the more diverse perspectives as to why natural phenomena occur on Planet Kazar, the more information available to consensus members to collectively achieve consensus accuracy. The developmental patterns of ratings across trials reflect this progression. That is, a *multiplist* understanding is consistent with children's increasing tendency to prefer diversity among consensus members, namely, a consensus with *distinct* group membership and *distant* proximity, and for that preference to weaken when there is *close* proximity between consensus members. Moreover, a multiplist understanding is also consistent with the tendency to refer to diversity as a mechanism for achieving consensus accuracy through 'Pooled Knowledge'. As a representative example of that more granular coding category, one child explained, "*they each go to a different school and so they each learn different things... you get more options and opinions of why the clouds are pink*."

Above and beyond a shift in directionality, an *evaluativist* understanding leads to patterns of ratings across trials that reflect a more nuanced valuing of the *distinct* group membership and *distant* relational properties. While the 6-point rating scale used in the current dissertation study

provided some insight as to the degree in variation among children's relative preferences across trials, a more sensitive measure is likely required to make any substantive conclusions. However, it is worth noting, adults' ratings across trials demonstrate more clearly the unique influence of *distinct* group membership and *distant* proximity relations (see Figure 16). In particular, if each cue uniquely informs judgments of relative preference, then we would expect that confidence in the consensus with *distinct* group membership would be at the highest level on the Aligned trial when there is *distant* proximity between them. Comparatively, on the Crossed trial, when *close* proximity undermines preferences for that consensus, preference ratings should indicate lower levels of confidence in the consensus with *shared* group membership.

With respect to the progression from multiplist to evaluativist levels of understanding, developmental trends in the proportion of *Mechanism* explanations that referenced 'Pooled Knowledge' and 'Consilience' may provide a starting point for further investigation. Specifically, although the overall frequency of *Mechanism* explanations increased with age , the proportion of explanations that referenced 'Pooled Knowledge' within this theme decreased with age, while those that referenced 'Consilience' increased (see Figure 17 in Chapter 4 Results).

Another point of consideration relates to the way that a multiplist understanding is typically framed in the literature on epistemological understanding. Specifically, the multiplist conception of knowledge as opinions of subjective reality is often referenced as correlated with a disinclination to think critically about whether a certain position has more merit than another (See Kuhn 2001). However, in the current study, children's preference for social diversity coupled with the tendency to explain how diversity generates epistemic value in justifying their choices, challenges this perspective. Children who provided explanations that referenced 'Pooled Knowledge' may very well have an understanding that individual's beliefs reflect their opinions,

or interpretations of a subjective reality, but nevertheless seem to appreciate both the factual nature of knowledge about natural phenomena and the epistemic value of a wider (diverse) range of "opinions" to draw upon.

Future Directions

There are several findings in the current dissertation that warrant further investigation. The sections below are conceptually organized to address four directions for future research.

Expectations for Group Members' Shared Experience. First, future work could conduct a more in-depth investigation of children's preferences for different patterns of consensus based on social group membership alone. A wider range of proximity cues (e.g., consensus members sharing the same classroom, same neighborhood, living in the same apartment building, etc.) could be manipulated to allow for more in-depth analysis of the overarching hypothesis that children's consensus group preferences reflect expectations for group members' shared experience. If children's preferences for either shared or distinct group membership among consensus members remain consistent with their preferences for a wider range of close or distant proximity cues, this would further support the idea that social group membership acts as a marker of shared experience. In turn, the perspective that social categories serve an epistemic function in judgments of reliability can inspire a range of future lines of research.

Developmental Trends in Preferences for Diversity. Another direction for future work is to further investigate the developmental progression demonstrated by the current findings with a broader age range. The current study corroborated the finding from the preliminary studies indicating that it is not until 8 years that children reliably prefer diversity among consensus members. However, given that the preliminary studies (i.e., Study 1A) did not examine

children's reasoning, it remains an open question as to whether there is developmental change between ages 5 and 7 in the explanatory process underlying children's reported preferences.

Epistemological Understanding as Developmental Mechanism. Moreover, by extending this work with a broader age range, future studies can directly examine the relation between children's epistemological understanding and their evaluations of consensus' accuracy. Again, manipulating a wider range of indirect cues to dependencies among consensus members would allow for a systematic investigation of epistemological understanding as a developmental mechanism underlying the ability to appreciate the epistemic value of social diversity.

Studies that have examined the development of epistemological understanding — from absolutist, to multiplist, and finally to evaluativist level—suggest that while the developmental sequence is consistent across domains, advancement in conceptual understanding in one domain does not necessarily transfer to others (e.g., Iordanou, 2010; 2016; Hofer, 2004). Future studies could build on the preliminary studies by leveraging the paradigm developed for the current dissertation in order to investigate whether there are domain-specific differences in children's patterns of consensus group preferences when *close* versus *distant* cues to proximity are present.

Real World Social Groups. The paradigm developed for the current dissertation can also be adapted in future work using stimuli representing actual human children belonging to real social categories, rather than novel social categories on a novel planet. This is suggested as a future direction, rather than a limitation of the current dissertation study, due to the theoretical perspective that has guided this work. Of course, replication of the findings in the current dissertation is certainly necessary for drawing strong conclusions about their reliability. Nevertheless, if children's preferences deviate from the patterns demonstrated in the current research, it is more likely a result of their representation of familiar social categories than that of

a different cognitive process underlying their choices in the current study. Many of the children (and almost all of the adults) in the current research spontaneously drew analogical comparisons between the social compositions of consensus groups comprised of aliens living on Planet Kazar to real world inter-group relations. The following three quotes are examples of these comparisons: (1) "...I'm guessing that's like a private school and those ones are like a public school... private schools are only for certain kinds of people." (2) "...In my school right now, we are learning about mountains. But another school that my friend goes to they are learning something else... So maybe their teachers are teaching them different things that might help..." (3) "...Christians believe in God. And other people don't. Because they are different, they will think different things, unless they are right."

Limitations

The current study has several limitations. First, participants only completed a single trial for each trial type. The decision to limit the number of trials was made in order to maintain participants' attention and motivation throughout the study session. While there are inherent benefits to additional trials and repeated measures for statistical analysis, the quality of data collected was prioritized over the quantity. Of relevance to this point, when asked to provide explanations on the latter trials, several of the 9-year-old participants answered, "same as last time." The experimenter requested that they kindly explain again, and the children proceeded. However, this signifies the challenges considered when deciding whether to include additional trials.

While the final sample was diverse in terms of geographic locations, the majority of child participants were White, monolingual English speakers, with at least one parent with a postgraduate degree. To provide more descriptive information about children's contexts, school-

level demographics were obtained, and RDI scores were computed to indicate the level of diversity or variation in a school's student population across different racial groups. The distribution of RDI scores across the age groups indicate a relatively wide range of scores, given the nature of school demographic patterns across schools in the United States. Nevertheless, the RDI does not provide insight beyond racial demographic diversity. The implications of the current study are therefore limited by the sample of children that participated.

In addition, study sessions with children were conducted online over Zoom. Conducting this research online allowed for reaching participants beyond the local geographic area, and enabled participation without the need to travel to a study site. However, there are inherent limitations to online research, including time lag in the presentation of stimuli, technical issues, and an inability for the experimenter to confirm participants' line of view. While no major issues impeding the completion of study sessions were reported, it is still possible that potential participants met barriers that prevented their ability to participate in online research.

Implications & Applications

The current work contributes to advancing theory in the trust and testimony developmental literature and children's developing Theory of Mind. Highlighted throughout chapters 1 and 2 of this dissertation, the influence of social group membership on children's decisions of whom and what to trust is commonly studied through the lens of children's own ingroup and outgroups. While this approach certainly provides important insights on how ingroup biases emerge and their impact on children's judgments of reliability, a broader perspective on how children come to represent the epistemic implications of patterns in the social world is critical. In daily life, children and adults alike often have no information about unfamiliar informants other than who they are and with whom they affiliate. The ability to not

only track others' relationships, but also reason about the implications of overlap in social networks, facilitates an ability to recognize potential shared sources of judgement errors – namely those that are systematic or biased toward a particular direction. Thus, how children integrate information across multiple sources is crucial for effectively evaluating the reliability of converging views.

Few studies have examined how dependencies between consensus members, particularly in situations where there are an equal number of endorsements for different positions, influence children's judgements of consensus reliability. However, of those few studies, none have directly examined children's representations of social diversity and how they may reason about its epistemic value. This dissertation study provides new insights as to how children account for the social relations of informants when evaluating their credibility. Previous research focusing on children's use of social relations has generally focused on subjective contexts and children's ability to distinguish between informants that are more or less likely to make judgments influenced by pre-existing biases. For example, studies have examined children's ability to recognize that people's judgements may be skewed by their relationships when their task is to judge contests in which the criteria is subjective (e.g., a dance contest) rather than objective (e.g., running a race; Mills & Grant, 2009; Mills & Keil, 2005, 2008). The current dissertation examined how children account for the social relations between consensus members in an objective context, where the task was learning the "right things" about why phenomena occur in the natural world. Importantly, the developmental trends demonstrated in the current work indicating that by 8-years of age children recognize that individuals' social relations can influence their judgements in an objective domain, is consistent with findings from that literature.

Children are growing up in a society where the ability to publish and circulate any information is often unchecked. It is not particularly challenging to identify sources with converging viewpoints on any given topic. The more critical challenge, therefore, is how to assess sources' reliability. Considering the relations between sources is a strategy for disentangling the influence of different perspectives on the same event. For example, an individual may watch coverage of a particular news story across various media outlets and find that the coverage is consistent across many, but not all. At face value, it is reasonable to assume that their inclination may be to trust the view of the majority of media outlets. However, the individual may learn that all of the converging reports originate from sources that share common ownership. In other words, the media outlets' converging reports share a common background, shaping the perspectives shared in the news coverage. In this case, adopting the view of the majority of media outlets would reflect an over-weighting of the converging reports' evidential value. Whereas, if the individual turned to the minority outlets, it would suggest that they accounted for redundant information introduced through sharing a common background (reducing evidential support). Indeed, fact checkers use such a strategy, referred to as "lateral reading" (e.g., Breakstone et al., 2021). Rather than examining the content of a single source, they search across multiple sources on the same issue to verify claims. The degree of consensus across sources is only as informative as the range of distinct perspectives represented.

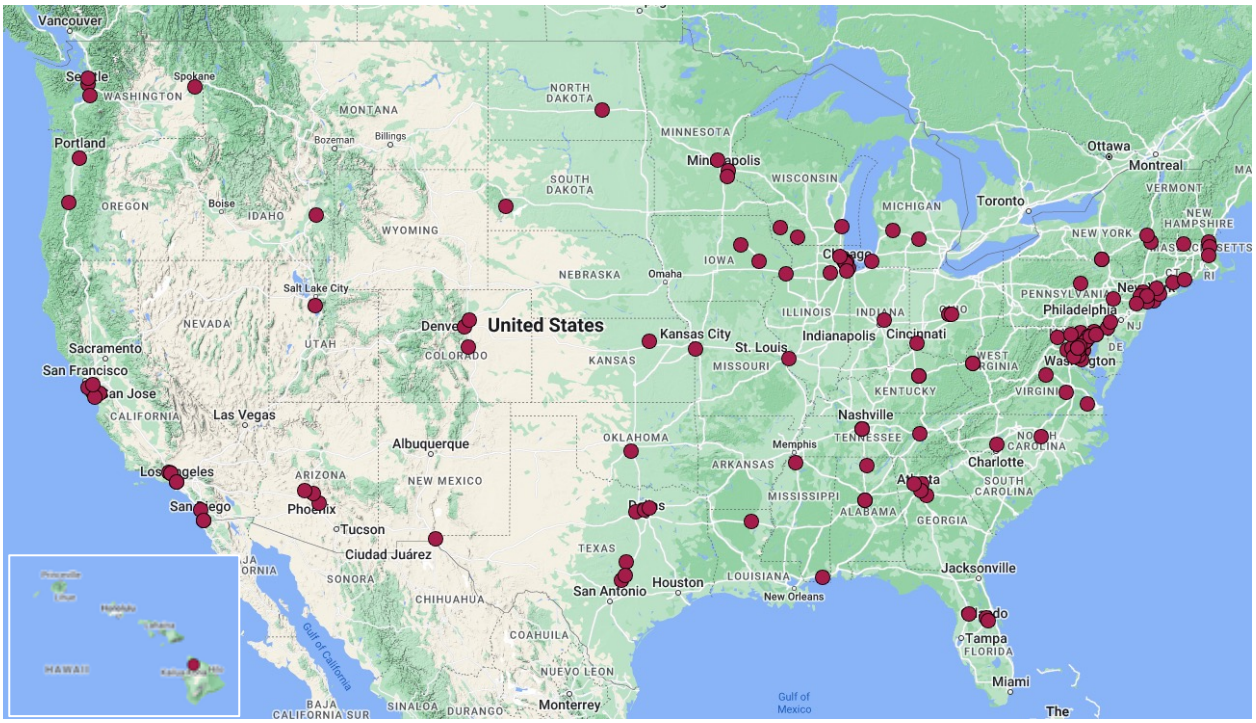
Conclusion

We all rely on consensus to build on the expertise of others, but certain consensus patterns provide better evidence for the truth of a given claim. The individuals comprising a consensus represent perspectives that are shaped by the information that flows through their social networks. This dissertation provides preliminary evidence that children develop an

increasing appreciation for the implications of social relations between consensus members. Different reasoning can underlie the same patterns of preferences for diversity among consensus members. What comes with an increasing preference for seeking the view endorsed by a diverse consensus seems to be a sense of how social diversity generates epistemic value. An ability to recognize cues to *indirect* dependence between sources introduced through common backgrounds is an essential foundation for determining the range of distinct perspectives represented on a given issue and is an important direction for future research.

Appendices

Appendix A: Map of Child Participant Geographic Locations



Note. Participation was limited to participants living in the United States (*n*=144)

Appendix B: Parent Demographic Survey Items

Please tell us a little about you and your child.
We greatly appreciate your time in providing this background information!

***Information provided will be kept confidential and participation does not depend on your answers to these questions. We only want to learn more about the children who participate in our research as a group.*

1. Child's birthday (MM/DD/YYYY):_____

2. Child's gender:
   - ☐ Non-Binary
   - ☐ Male
   - ☐ Female
   - ☐ Prefer not to answer
   - ☐ Other_____

3. Child's Race:
   - ☐ Asian
   - ☐ Black/African
   - ☐ Hispanic/Latinx
   - ☐ Native American
   - ☐ Pacific Islander
   - ☐ White
   - ☐ Other_____

4. Language(s) spoken by your child:
   - ☐ English
   - ☐ Other (please specify):_____
   - ☐ Other (please specify):_____
   - ☐ Other (please specify):_____

5. The zip code in which your child lives: _____

6. Name of the school which your child attends: _____

7. Please indicate the highest level of education completed by each of the child's parents:

Parent 1:

- Some high school
- High School Diploma/GED
- Some College Coursework/Vocational Training
- 2-year College Degree (Associates)
- 4-year College Degree (BA/BS)
- Postgraduate or Professional degree (MA, PhD, MD, JD)
- Prefer not to answer

Parent 2:

- Some high school
- High School Diploma/GED
- Some College Coursework/Vocational Training
- 2-year College Degree (Associates)
- 4-year College Degree (BA/BS)
- Postgraduate or Professional degree (MA, PhD, MD, JD)
- Prefer not to answer
- Not Applicable

8. Please indicate your annual household income:

- Less than $15,000
- $15,000 - $30,000
- $31,000 - $45,000
- $46,000 - $59,000
- $60,000 - $75,000
- $76,000 - $100,000
- $101,000 - $150,000
- $151,000 or more
- Prefer not to answer

9. Is there anything else you would like to know? _____

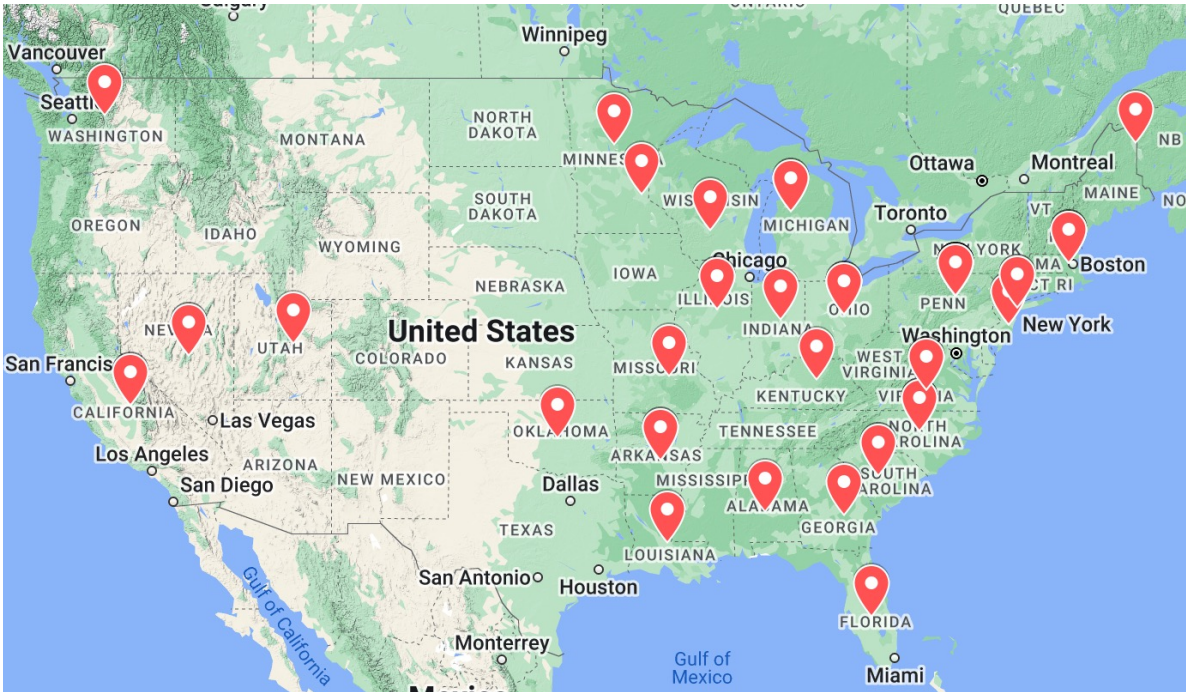
Appendix C: Descriptive Statistics of Child Demographic Survey Variables

	7-year-olds *n* = 48	8-year-olds *n* = 48	9-year-olds *n*=48
Age in Years, *M* (*SD*)	7.5 (0.27)	8.47 (0.31)	9.42 (0.3)
[min, max]	[7, 7.9]	[8, 8.9]	[9, 9.9]
Child Gender, *n* (*%*)			
Male	24 (50)	24 (50)	24 (50)
Female	24 (50)	24 (50)	24 (50)
Child Race, *n* (*%*)			
Asian	7 (14.6)	9 (18.8)	4 (8.3)
Biracial/Multiracial	9 (18.8)	7 (14.6)	11 (22.9)
Black/African	1 (2.1)	1 (2.1)	2 (4.2)
Hispanic/Latinx	3 (6.3)	1 (2.1)	2 (4.2)
White	27 (56.3)	30 (62.5)	27 (56.3)
Other[a]	1 (2.1)	0(0)	1 (2.1)
Language, *n* (%)			
Monolingual	38 (79.2)	42 (87.5)	39 (81.3)
Bilingual[b]	8 (16.7)	6 (12.5)	6 (12.5)
Trilingual[c]	2 (4.2)	0 (0)	3 (6.3)
Parent 1 Education, *n* (*%*)			
High School Diploma/GED	0 (0)	1 (2.1)	1 (2.1)
Some College Coursework/Vocational Training	4 (8.3)	2 (4.2)	2 (4.2)
2-year College Degree (Associates)	1 (2.1)	0 (0)	0 (0)
4-year College Degree (BA/BS)	19 (39.6)	8 (16.7)	16 (33.3)
Postgraduate or Professional Degree (MA, PhD, MD, JD)	23 (47.9)	37 (77.1)	28 (58.3)
Some high school	1 (2.1)	0 (0)	0 (0)
Parent 2 Education, *n* (*%*)			
High School Diploma/GED	0 (0)	1 (2.1)	0 (0)
Some College Coursework/Vocational Training	4 (8.3)	1 (2.1)	3 (6.3)
2-year College Degree (Associates)	3 (6.3)	3 (6.3)	2 (4.2)
4-year College Degree (BA/BS)	16 (33.3)	12 (25.0)	14 (29.2)
Postgraduate or Professional Degree (MA, PhD, MD, JD)	23 (47.9)	28 (58.3)	27 (56.3)
School Type, *n* (%)			
Public School	31 (64.6)	29 (60.4)	36 (75.0)
Charter School	3 (6.3)	3 (6.3)	0 (0)
Private School (Non-Religious)	1 (2.1)	1 (2.1)	2 (4.2)
Private (Religious)	4 (8.3)	5 (10.4)	5 (10.4)
Homeschool	6 (12.5)	5 (10.4)	4 (8.3)
Not Reported	3 (6.3)	5 (0)	1 (2.1)

Annual Household Income, *n* (%)			
Less than $15,000	2 (4.2)	0 (0)	0 (0)
$15,000 - $30,000	0 (0)	0 (0)	2 (4.2)
$31,000 - $45,000	1 (2.1)	2 (4.2)	0 (0)
$46,000 - $59,000	3 (6.3)	5 (.1)	5 (.1)
$60,000 - $75,000	6 (12.5)	2 (4.2)	4 (8.3)
$76,000 - $100,000	4 (8.3)	4 (8.3)	7 (14.6)
$101,000 - $151,000	15 (31.3)	8 (16.6)	10 (20.8)
$151,000 or more	8 (16.6)	16 (33.3)	10 (20.8)
Not Reported	9 (18.8)	11 (22.9)	9 (18.8)

[a]"Other" races included: North African.
[b]"Bilingual" languages included: Brushaski, Chinese, Dutch, French, Malayalam, Mandarin, Spanish, Urdu
[c]"Trilingual" languages included: Hindi, Italian, Mandarin, Spanish, Russian, German

Appendix D: Map of Adult Participant Geographic Locations



Note. Participation was limited to participants living in the United States (n=48).

Appendix E: Descriptive Statistics of Prolific Demographics for Adult Sample

	Adults (n = 48)
Age in Years, *M* (*SD*)	37.65 (11.84)
[min, max]	[19, 61]
Adult Gender, *n* (%)	
Male	26 (54.17)
Female	21 (43.75)
Non-Binary	1 (2.08)
Race/Ethnicity, *n* (%)	
Black	2 (4.17)
White	38 (79.17)
Asian	4 (8.33)
Mixed	3 (11.5)
Other	1 (2.08)
Highest Level of Education, *n* (%)	
High school diploma/A-levels	11 (22.92)
Technical/community college	7 (14.58)
Undergraduate degree (BA/BSc/other)	20 (41.67)
Graduate degree (MA/MSc/MPhil/other)	6 (12.50)
Doctorate degree (PhD/other)	4 (8.33)
Employment, *n* (%)	
Full-Time	23 (47.92)
Part-Time	13 (27.08)
Unemployed (and job seeking)	5 (10.42)
Not in Paid Work (e.g., 'homemaker', retired or disabled)	6 (12.50)
Other	1 (2.08)
Annual Household Income, *n* (%)	
Less than $10,000	1 (2.08)
$20,000 - $29,999	8 (16.17)
$30,000 - $39,999	4 (8.33)
$40,000 - $49,999	8 (16.17)
$50,000 - $59,999	3 (6.25)
$60,000 - $69,999	1 (2.08)
$70,000 - $79,999	2 (4.17)
$80,000 - $89,999	1 (2.08)

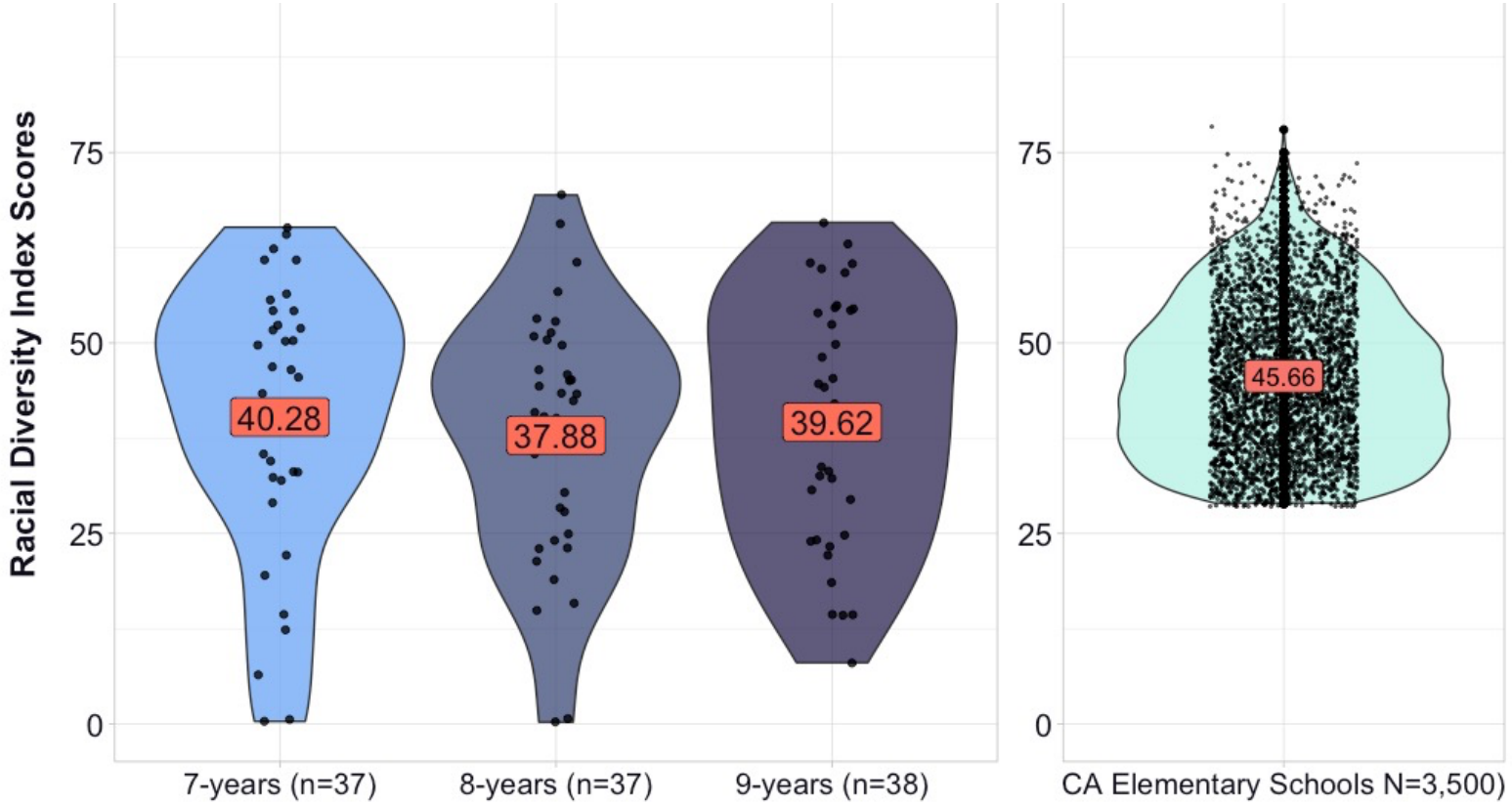
$90,000 - $99,999	4 (8.33)
$100,000 - $150,000	10 (20.83)
$151,000 or more	6 (12.50)

Appendix F: Descriptive Statistics of Racial Diversity Index Scores

	7-year-olds *n* = 48	8-year-olds *n* = 48	9-year-olds *n*=48
Total Sample			
School Reported	37 (77.1)	37 (77.1)	38 (79.2)
School Not Reported	2 (4.2)	4 (8.3)	0 (0)
School Demographics Not Reported[a]	9 (18.8)	7 (14.6)	10 (20.8)
Ethnic Diversity Index, *M* (*SD*)	40.3 (17.6)	37.9 (16.4)	39.6 (16.0)
[min, max]	[.34, 65.19]	[.26, 69.43]	[8.04, 65.83]

[a] "School Demographics Not Reported" indicates the private schools for which no demographic information are reported.

Appendix G: Distribution of School-Level Demographics Index Scores



Note. Violin plots and means across age groups (left panel) and for all public elementary schools in California (CA) during the academic year of 2021-2022 as a comparison group (right panel). Distribution of Racial Diversity Index (RDI) scores (y axis), ranging from 0 to 100, with higher scores indicating a more even distribution of number of students across total racial categories reported. Mean scores across are in the red boxes.

Appendix H: IRB Approval Letter

UNIVERSITY OF MARYLAND
INSTITUTIONAL REVIEW BOARD

1204 Marie Mount Hall
College Park, MD 20742-5125
TEL 301.405.4212
FAX 301.314.1475
irb@umd.edu
www.umresearch.umd.edu/IRB

DATE:	March 1, 2023
TO:	Lucas Butler, PhD
FROM:	University of Maryland College Park (UMCP) IRB
PROJECT TITLE:	[1470042-10] Reasoning About the Reliability of Consensus Testimony
SUBMISSION TYPE:	Amendment/Modification
ACTION:	APPROVED
APPROVAL DATE:	March 1, 2023
REVIEW TYPE:	Expedited Review
REVIEW CATEGORY:	Expedited review category #7. 45CFR46.404 applies. 45CFR46.404 applies. Waiver of consent documentation (online verbal consent & adult follow up) - 45CFR46.117(c). Alteration of parental consent (opt out) - 45CFR46.116(f)/45CFR46.408(c).

Thank you for your submission of Amendment/Modification materials for this project. The University of Maryland College Park (UMCP) IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a project design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

Prior to final approval of this project scientific review was completed by the IRB Member reviewer.

This submission has received Expedited Review based on the applicable federal regulations.

This project has been determined to be a MINIMAL RISK project.

Please remember that informed consent is a process beginning with a description of the project and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the project via a dialogue between the researcher and research participant. Unless a consent waiver or alteration has been approved, Federal regulations require that each participant receives a copy of the consent document.

Please note that any revision to previously approved materials must be approved by this committee prior to initiation. Please use the appropriate Amendment forms for this procedure.

All UNANTICIPATED PROBLEMS involving risks to subjects or others (UPIRSOs) and SERIOUS and UNEXPECTED adverse events must be reported promptly to this office. Please use the appropriate reporting forms for this procedure. All FDA and sponsor reporting requirements should also be followed. All NON-COMPLIANCE issues or COMPLAINTS regarding this project must be reported promptly to this office.

Please note that all research records must be retained for a minimum of seven years after the completion of the project.

If you have any questions, please contact the IRB Office at 301-405-4212 or irb@umd.edu. Please include your project title and reference number in all correspondence with this committee.

This letter has been electronically signed in accordance with all applicable regulations, and a copy is retained within University of Maryland College Park (UMCP) IRB's records.

References

Aboody, R., Yousif, S. R., Sheskin, M., & Keil, F. C. (2022). Says who? Children consider informants' sources when deciding whom to believe. *Journal of Experimental Psychology: General*, *151*(10), 2481–2493. https://doi.org/10.1037/xge0001198

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition, 21*(1), 37-46.

Birch, S. A., Akmal, N., & Frampton, K. L. (2010). Two-year-olds are vigilant of others' non-verbal cues to credibility. *Developmental Science, 13*(2), 363–369. doi:10.1111/j.1467-7687.2009.00906.x

Birch, S. A. J., Vauthier, S. A., & Bloom, P. (2008). Three- and four-year olds spontaneously use others' past performance to guide their learning. *Cognition, 107*, 1018 –1034. https://doi.org/10.1016/j.cognition.2007.12.008

Blakemore, J. E. O. (2003). Children's beliefs about violating gender norms: Boys shouldn't look like girls, and girls shouldn't act like boys. *Sex Roles, 48*(9-10), 411–419. doi:10.1023/A:1023574427720

Bond, R. (2005). Group size and conformity. *Group Processes & Intergroup Relations, 8*(4), 331–354.

Breakstone, J., Smith, M., Connors, P., Ortega, T., Kerr, D., & Wineburg, S. (2021). Lateral reading: College students learn to critically evaluate internet sources in an online course. *Harvard Kennedy School Misinformation Review, 2*(1), 1-17. https://doi.org/10.37016/mr-2020-56

Brosseau-Liard, P.E. & Birch, S.A.J. (2011). Epistemic states and traits: Preschoolers appreciate the differential informativeness of situation-specific and person-specific cues to knowledge. *Child Development, 82,* 1788-1796. doi: 10.1111/j.1467-8624.2011.01662.x.

Burdett, E. R. R., Lucas, A. J., Buchsbaum, D., McGuigan, N., Wood, L. A. & Whiten, A. (2016). Do children copy an expert or a majority? Examining selective learning in instrumental and normative contexts. *PLoS ONE, 11*(10), e0164698.

Buresh, J. S. & Woodward, A. L. (2007). Infants track action goals within and across agents. *Cognition, 104*, 287–314.

Butler, L. P., Gibbs, H. M., & Tavassolie, N. S. (2020). Children's developing understanding that even reliable sources need to verify their claims. *Cognitive Development, 54*, 100871.

Butler, L. P., Schmidt, M. F. H., Tavassolie, N., & Gibbs, H. (2018). Children's evaluation of verified and unverified claims. *Journal of Experimental Child Psychology, 176*, 73-83. doi: 10.1016/j.jecp.2018.07.007

Carey, S. Conceptual change in childhood. Cambridge, Mass: MIT Press; 1985.

Carey, S. (1988). Conceptual differences between children and adults. *Mind and Language, 3*(3), 167-181.

Carpendale, J. I., & Chandler, M. J. (1996). On the distinction between false belief understanding and subscribing to an interpretive theory of mind. *Child Development, 67*(4), 1686-1706. https://doi.org/10.1111/j.1467-8624.1996.tb01821.x

Chandler, M. J., & Carpendale, J. I. M. (1998). Inching toward a mature theory of mind. In M. D. Ferrari & R. J. Sternberg (Eds.), *Self-awareness: Its nature and development* (pp. 148–190). Guilford Press.

Caza, J. S., Atance, C. M., & Bernstein, D. M. (2016). Older (but not younger) preschoolers understand that knowledge differs between people and across time. *British Journal of Developmental Psychology, 34*(3), 313-324.

Chandler, M. (1988). Doubt and developing theories of mind. In J. Astington, P. Harris, & D. Olson (Eds.), *Developing theories of mind* (pp. 387-413). New York: Cambridge University Press.

Chinn, C. A., Buckland, L. A., & Samarapungavan, A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. *Educational Psychologist,46*(3), 141-167.doi: http://dx.doi.org/10.1080/00461520.2011.587722

Chignell, A. (2010). The ethics of belief. In *The Stanford Encyclopedia of Philosophy (Summer 2010 Edition*). Retrieved from http://plato.stanford.edu/entries/ethics-belief/

Cimpian, A., & Scott, R. M. (2012). Children expect generic knowledge to be widely shared. *Cognition,123*(3), 419-433.

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

Coady, C. A. J. (1992). *Testimony: A Philosophical Study.* Oxford: Oxford University Press

Corriveau, K. H., Kinzler, K. D., & Harris, P. L. (2013). Accuracy trumps accent in children's endorsement of object labels. *Developmental Psychology, 49*(3), 470-479. doi: 10.1037/a0030604

Corriveau, K. H., Fusaro, M., & Harris, P. L. (2009). Going with the flow: Preschoolers prefer nondissenters as informants. *Psychological Science, 20*(3), 372-377.

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

Denison, S., & Xu, F. (2010). Integrating physical constraints in statistical inference by 11-month-old infants. *Cognitive Science, 34*(5), 885-908. https://doi.org/10.1111/j.1551-6709.2010.01111.x

Doebel, S., Rowell, S. F., & Koenig, M. A. (2016). Young children detect and avoid logically inconsistent sources: The importance of communicative context and executive function. *Child Development, 87*(6), 1956-1970.

Einav, S. (2018). Thinking for themselves? The effect of informant independence on children's endorsement of testimony from a consensus. *Social Development, 27*(1), 73-86.

Einav, S., & Robinson, E. J. (2011). When being right is not enough: Four-year-olds distinguish knowledgeable informants from merely accurate informants. *Psychological Science*, *22*(10), 1250-1253. doi: 10.1177/0956797611416998

Eisbach, A. O. D. (2004). Children's developing awareness of diversity in people's trains of thought. *Child Development, 75*(6), 1694-1707.

Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*(6), 878. https://doi.org/10.1037/0022-3514.82.6.878

Fitneva, S. A. (2008). The role of evidentiality in Bulgarian children's reliability judgments. *Journal of Child Language, 35*, 845–868. doi:10.1017/S0305000908008799

Fitneva, S. A. (2010). Children's representation of child and adult knowledge. *Journal of Cognition and Development, 11*(4), 458-484.

Fitneva, S. A., Lam, N. H., & Dunfield, K. A. (2013). The development of children's information gathering: To look or to ask? *Developmental Psychology, 49*(3), 533.

Gelman, S. A., Ware, E. A., & Kleinberg, F. (2010). Effects of generic language on category content and structure. *Cognitive Psychology, 61*(3), 273–301. doi:10.1016/j.cogpsych.2010.06.001

Gelman, S. A., Ware, E. A., & Kleinberg, F. (2010). Effects of generic language on category content and structure. *Cognitive Psychology, 61*(3), 273-301. https://doi.org/10.1016/j.cogpsych.2010.06.001

Gopnik, A. (1983). Conceptual and semantic change in scientists and children: why there are no semantic universals. *Linguistics*, *21*(1), 163-180. https://doi.org/10.1515/ling.1983.21.1.163

Gopnik, A. (1984). The acquisition of gone and the development of the object concept. *Journal of Child Language, 11*(2), 273-292.

Gopnik, A. (1988). Conceptual and semantic development as theory change: The case of object permanence. *Mind & Language*, 3(3), 197–216. doi:10.1111/j.1468-0017.1988.tb00143.x

Gopnik, A. & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development, 59*, 26-37.

Gopnik, A., & Graf, P. (1988). Knowing how you know: Young children's ability to identify and remember the sources of their beliefs. *Child Development*, 1366-1371.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A Theory of Causal Learning in Children: Causal Maps and Bayes Nets. *Psychological Review, 111*(1), 3–32. https://doi.org/10.1037/0033-295X.111.1.3

Gopnik, A. & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind*

& Language, 7, 145–171. doi:10.1111/j.1468-0017.1992.tb00202.x

Gopnik, A., & Wellman, H. M. (1994). The theory theory. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 257–293). Cambridge University Press. https://doi.org/10.1017/CBO9780511752902.011

Gutheil, G., & Gelman, S. A. (1997). Children's use of sample size and diversity information within basic-level categories. Journal of Experimental Child Psychology, 64(2), 159-174. https://doi.org/10.1006/jecp.1996.2344

Hahn, U., Harris, A. J., & Corner, A. (2016). Public reception of climate science: Coherence, reliability, and independence. *Topics in Cognitive Science*, *8*(1), 180-195. doi:10.1111/tops.12173

Harris, P. L., & Koenig, M. A. (2006). Trust in testimony: How children learn about science and religion. *Child Development, 77*(3), 505-524.

Heiphetz, L., Spelke, E. S., Harris, P. L., & Banaji, M. R. (2013). The development of reasoning about beliefs: Fact, preference, and ideology. *Journal of Experimental Social Psychology, 49*(3), 559-565. https://doi.org/10.1016/j.jesp.2012.09.005

Hu, J. C., Whalen, A., Buchsbaum, D., Griffiths, T. L., & Xu, F. (2015). Can children balance the size of a majority with the quality of their information? In R. Dale, C. Jennings, P. P. Maglio, T. Matlock, D. C. Noelle, A. Warlaumont, J. Yoshimi (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 956-961). Merced, CA: Cognitive Science Society.

Huh, M., Grossmann, I., & Friedman, O. (2019). Children show reduced trust in confident advisors who are partially informed. *Cognitive Development, 50*, 49-55.

Iordanou, K. (2016). From theory of mind to epistemic cognition. A lifespan perspective.

Frontline Learning Research, 4(5), 106-119. https://doi.org/10.14786/flr.v4i5.252

Jaswal, V. K., & Malone, L. S. (2007). Turning believers into skeptics: 3-Year-olds' sensitivity to cues to speaker credibility. *Journal of Cognition & Development, 8*(3), 263–283. doi:10.1080/15248370701446392

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*(9), 757–758. https://doi.org/10.1111/j.1467-9280.2006.01778.x

Karadağ, D., & Soley, G. (2023). Children intend to teach conventional but not moral norms selectively to ingroup members. *Developmental Psychology, 59*(3), 567–578. https://doi.org/10.1037/dev0001455

Karmiloff-Smith, A., & Inhelder, B. (1974). If you want to get ahead, get a theory. *Cognition, 3*(3), 195-212.

Karmiloff-Smith, A. (1988). The child is a theoretician, not an inductivist. *Mind & Language, 3*(3), 183-196.

Keil, F. C. (1989). Concepts, kinds, and cognitive development. Cambridge, Mass: MIT Press.

Kim, S., & Spelke, E. S. (2020). Learning from multiple informants: Children's response to epistemic bases for consensus judgments. *Journal of Experimental Child Psychology, 192,* 104759.

Kinzler, K. D., Corriveau, K. H., & Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Developmental Science, 14*(1), 106-111. doi: 10.1111/j.1467-7687.2010.00965.x

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences, 104*(30), 12577-12580.

Koenig, M. A. (2012). Beyond semantic accuracy: Preschoolers evaluate a speaker's reasons. *Child Development, 83*(3), 1051-1063. https://doi.org/10.1111/j.1467-8624.2012.01742.x

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development, 76*, 1261–1277. doi:10.1111/j.1467-8624.2005.0084

Koenig, M., Clément, M., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science, 15,* 694 – 698. doi: 10.1111/j.0956-7976.2004.00742.x

Koenig, M. A., & Jaswal, V. K. (2011). Characterizing children's expectations about expertise and incompetence: Halo or pitchfork effects?. *Child Development, 82*(5), 1634-1647. https://doi.org/10.1111/j.1467-8624.2011.01618.x

Koenig, M. A., & Woodward, A. L. (2010). Sensitivity of 24-month-olds to the prior inaccuracy of the source: possible mechanisms. *Developmental psychology*, *46*(4), 815. https://doi.org/10.1037/a0019664

Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Science, 9*(5), 178-181.

Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development, 15*(3), 309-328.

Kushnir, T., Vredenburgh, C., & Schneider, L. A. (2013). "Who can help me fix this toy?" The distinction between causal knowledge and word knowledge guides preschoolers' selective requests for information. *Developmental Psychology, 49*, 446–453. doi:10.1037/a0031649

Lalonde, C. E., & Chandler, M. J. (2002). Children's understanding of interpretation. *New Ideas in Psychology, 20*(2-3), 163-198. https://doi.org/10.1016/S0732-118X(02)00007-7

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: 10.1111/cdev.12324

Lane, J. D., Ronfard, S. & El-Sherif, D. (2018). The influence of first-hand testimony and hearsay on children's belief in the improbable. *Child Development, 89*(4), 1133-1140. do 10.1111/cdev.12815

Liben, L. S., Bigler, R. S., & Krogh, H. R. (2002). Language at work: Children's gendered interpretations of occupational titles. *Child Development, 73*(3), 810–828. https://doi.org/10.1111/1467-8624.00440

Liberman, Z., Gerdin, E., Kinzler, K. D., & Shaw, A. (2020). (Un) common knowledge: Children use social relationships to determine who knows what. *Developmental Science, 23*(6), e12962.

Liberman, Z., Kinzler, K. D., & Woodward, A. L. (2018). The early social significance of shared ritual actions. *Cognition, 171*, 42-51.

Liberman, Z., & Shaw, A. (2017). Children use partial resource sharing as a cue to friendship. *Journal of Experimental Child Psychology, 159*, 96-109.

Liberman, Z., & Shaw, A. (2018). Secret to friendship: Children make inferences about friendship based on secret sharing. *Developmental Psychology, 54*(11), 2139.

Liberman, Z., & Shaw, A. (2019). Children use similarity, propinquity, and loyalty to predict which people are friends. *Journal of Experimental Child Psychology, 184*, 1-17.

Liberman, Z., & Shaw, A. (2020). Even his friend said he's bad: Children think personal alliances bias gossip. *Cognition, 204*, 104376.

Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What

could possibly go wrong?. *Journal of Experimental Social Psychology, 79*, 328-348. https://doi.org/10.1016/j.jesp.2018.08.009

Lockhart, K. L., Goddu, M. K., Smith, E. D., & Keil, F. C. (2016). What could you really learn on your own?: Understanding the epistemic limitations of knowledge acquisition. *Child Development, 87*(2), 477-493.

Lutz, D. J., & Keil, F. C. (2002). Early understanding of the division of cognitive labor. *Child Development, 73*, 1073 – 1084. https://doi.org/10.1111/1467-8624.00458

Marazita, J. M., & Merriman, W. E. (2004). Young children's judgment of whether they know names for objects: The metalinguistic ability it reflects and the processes it involves. *Journal of Memory and Language, 51*(3), 458–472. https://doi.org/10.1016/j.jml.2004.06.008

Marble, K. E., & Boseovski, J. J. (2019). Children's judgments of cultural expertise: The influence of cultural status and learning method. *The Journal of Genetic Psychology, 180*(1), 17-30. https://doi.org/10.1080/00221325.2018.1562418

Matsui, T., Yamamoto, T., & McCagg, P. (2006). On the role of language in children's early understanding of others as epistemic beings. *Cognitive Development, 21*,158–173. doi:10.1016/j.cogdev.2005.10.001

Miller, S. A., Hardin, C. A., & Montgomery, D. E. (2003). Young children's understanding of the conditions for knowledge acquisition. *Journal of Cognition & Development, 4*(3), 325-356.

Mills, C. M., & Keil, F. C. (2008). Children's developing notions of (im) partiality. *Cognition, 107*(2), 528-551. https://doi.org/10.1016/j.cognition.2007.11.003

Mills, C. M., & Landrum, A. R. (2016). Learning who knows what: Children adjust their inquiry

to gather information from others. *Frontiers in Psychology, 7,* 951. https://doi.org/10.3389/fpsyg.2016.00951

Morgan, T. J. H., Laland, K. N. & Harris, P. L. (2015) The development of adaptive conformity in young children: effects of uncertainty and consensus. *Developmental Science, 18*(4), 511–524

Morgan, T. J. H., Rendell, L. E., Ehn, M., Hoppitt, W. & Laland, K. N. (2012). The evolutionary basis of human social learning. *Proceedings of the Royal Society of London B: Biological Sciences, 279*(1729), 653–662.

Nurmsoo, E., & Robinson, E. J. (2009). Children's trust in previously inaccurate informants who were well or poorly informed: When past errors can be excused. *Child Development, 80*(1), 23-27. https://doi.org/10.1111/j.1467-8624.2008.01243.x

Park, S. A., Miller, D. S., & Boorman, E. D. (2021). Inferences on a multidimensional social hierarchy use a grid-like code. *Nature Neuroscience, 24*(9), 1292-1301. https://doi.org/10.1038/s41593-021-00

Pasquini, E., Corriveau, K., Koenig, M., & Harris, P. (2007). Children monitor the relative accuracy of informants. *Developmental Psychology, 43*, 1216 –1226. https://doi.org/10.1037/0012-1649.43.5.1216

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology, 5*(2), 125-137.

Perner, J., Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science, 308*(5719), 214–216. doi:10.1126/science.1111656

Piaget, J., & Cook, M. (1952). *The origins of intelligence in children (Vol. 8).* International

Universities Press.

Pillow, B. H. (1989). Early understanding of perception as a source of knowledge. *Journal of Experimental Child Psychology, 47*(1), 116-129. https://doi.org/10.1016/0022-0965(89)90066-0

Phillips, K., Hahn, U., & Pilditch, T. D. (2018). Evaluating testimony from multiple witnesses: single cue satisficing or integration? In T.T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 2244-2249). Austin, TX: Cognitive Science Society.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: Bradford Books/MIT Press.

Powell, L.J. & Spelke, E. S. (2018). Human infants' understanding of social imitation: Inferences of affiliation from third-party observations. *Cognition, 170*, 31–4832.

Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development, 61*(4), 973-982. https://doi.org/10.1111/j.1467-8624.1990.tb02835.x

Rhodes, M., & Brickman, D. (2010). The role of within-category variability in category-based induction: A developmental study. *Cognitive Science, 34*(8), 1561-1573. https://doi.org/10.1111/j.1551-6709.2010.01137.x

Rhodes, M., Brickman, D., & Gelman, S. A. (2008). Sample diversity and premise typicality in inductive reasoning: Evidence for developmental change. *Cognition, 108*(2)*,* 543–556. 10.1016/j.cognition.2008.03.002

Rhodes, M. & Chalik, L. (2013). Social categories as markers of intrinsic interpersonal obligations. *Psychological Science, 24*(6), 999–1006. doi:10.1177/0956797612466267

Roberts, S. O., & Gelman, S. A. (2015). Do children see in Black and White? Children's and adults' categorizations of multiracial individuals. *Child Development, 86*(6), 1830-1847. https://doi.org/10.1111/cdev.12410

Roberts, S. O., Gelman, S. A., & Ho, A. K. (2017). So it is, so it shall be: Group regularities license children's prescriptive judgments. *Cognitive Science, 41*, 576-600. https://doi.org/10.1111/cogs.12443

Robinson, E. J., Butterfill, S. A., & Nurmsoo, E. (2011). Gaining knowledge via other minds: Children's flexible trust in others as sources of information. *British Journal of Developmental Psychology, 29*(4), 961-980. https://doi.org/10.1111/j.2044-835X.2011.02036.x

Robinson, E. J., Champion, H., & Mitchell, P. (1999). Children's ability to infer utterance veractity from speaker informedness. *Developmental Psychology, 35*(2), 535-546.

Robinson, E. J., Haigh, S. N., & Nurmsoo, E. (2008). Children's working understanding of knowledge sources: Confidence in knowledge gained from testimony. *Cognitive Development, 23*(1), 105-118

Robinson, E. J., & Whitcombe, E. L. (2003). Children's suggestibility in relation to their understanding about sources of knowledge. *Child Development, 74*(1), 48-62.

Rohwer, M., Kloo, D., & Perner, J. (2012). Escape from metaignorance: How children develop an understanding of their own lack of knowledge. *Child Development, 83*(6), 1869-1883. https://doi.org/10.1111/j.1467-8624.2012.01830.x

Ross, H. S., Recchia, H. E., & Carpendale, J. I. (2005). Making sense of divergent interpretations of conflict and developing an interpretive understanding of mind. *Journal of Cognition and Development, 6*(4), 571-592. https://doi.org/10.1207/s15327647jcd0604_7

Ruffman, T. K., & Olson, D. R. (1989). Children's ascriptions of knowledge to others. *Developmental Psychology, 25*(4), 601. https://doi.org/10.1037/0012-1649.25.4.601

Scofield, J., & Behrend, D. (2008). Learning words from reliable and unreliable speakers. *Cognitive Development, 23*, 278 –290. https://doi.org/10.1016/j.cogdev.2008.01.003

Schulz, L. E., & Gopnik, A. (2004). Causal learning across domains. *Developmental Psychology, 40*(2), 162. https://doi.org/10.1016/j.tics.2004.06.005

Schulz, L. E., Gopnik, A., & Glymour, C. (2007). Preschool children learn about causal structure from conditional interventions. *Developmental Science, 10*(3), 322-332. https://doi.org/10.1111/j.1467-7687.2007.00587.x

Shafto, P., Eaves, B., Navarro, D. J., & Perfors, A. (2012). Epistemic trust: Modeling children's reasoning about others' knowledge and intent. *Developmental Science, 15*(3), 436-447. doi: 10.1111/j.1467-7687.2012.01135.x

Shtulman, A., & Carey, S. (2007). Improbable or impossible? How children reason about the possibility of extraordinary events. *Child Development, 78*(3), 1015-1032. https://doi.org/10.1111/j.1467-8624.2007.01047.x

Shutts, K., Banaji, M. R., & Spelke, E. S. (2010). Social categories guide young children's preferences for novel objects. *Developmental Science, 13*(4), 599-610.

Shutts, K., Pemberton Roben, C. K., & Spelke, E. S. (2013). Children's use of social categories in thinking about people and social relationships. *Journal of Cognition and Development, 14*(1), 35–62. doi:10.1080/ 15248372.2011.638686

Shutts, K., Roben, C. K. P., & Spelke, E. S. (2013). Children's use of social categories in thinking about people and social relationships. *Journal of Cognition and Development, 14*(1), 35-62. https://doi.org/10.1080/15248372.2011.638686

Sobel, D. M., & Corriveau, K. H. (2010). Children monitor individuals' expertise for word learning. *Child Development, 81*(2), 669-679. https://doi.org/10.1111/j.1467-8624.2009.01422.x

Sobel, D. M., & Munro, S. E. (2009). Domain generality and specificity in children's causal inference about ambiguous data. *Developmental Psychology, 45*(2), 511–524. https://doi.org/10.1037/a0014944

Sóley, G. (2019). What do group members share? The privileged status of cultural knowledge for children. *Cognitive Science, 43*(10), e12786.

Sóley, G., & Aldan, P. (2020). Children and adults selectively attribute shared cultural knowledge to speakers of the same language. *Child Development, 91*(1), e218-e230.

Sóley, G., & Köseler, B. (2021). The social meaning of common knowledge across development. *Cognition, 215*, 104811.

Stock, H. R., Graham, S. A., & Chambers, C. G. (2009). Generic language and speaker confidence guide preschoolers' inferences about novel animate kinds. *Developmental Psychology, 45*(3), 884–888. doi:10.1037/a0015359

Tardif, T., Wellman, H. M., Fung, K. Y. F., Liu, D., & Fang, F. (2005). Preschoolers' understanding of knowing- that and knowing- how in the United States and Hong Kong. *Developmental Psychology, 41*, 562– 573. doi:10.1037/0012-1649.41.3.562

Taylor, M. (1988). Conceptual perspective taking: Children's ability to distinguish what they know from what they see. *Child Development, 59,* 703–718. doi:10.2307/1130570

Taylor, M. G. (2013). Gender influences on children's selective trust of adult testimony. *Journal of Experimental Child Psychology, 115,* 672-690. doi:10.1016/j.jecp.2013.04.003

Terrier, N., Bernard, S., Mercier, H., & Clément, F. (2016). Visual access trumps gender in 3- and 4-year-old children's endorsement of testimony. *Journal of Experimental Child Psychology, 146,* 223-230. doi: 10.1016/j.jecp.2016.02.002.

Thomas, A. J., Woo, B., Nettle, D., Spelke, E., & Saxe, R. (2022). Early concepts of intimacy: young humans use saliva sharing to infer close relationships. *Science, 375*(6578), 311-315.

VanderBorght, M., & Jaswal, V. K. (2009). Who knows best? Preschoolers sometimes prefer child informants over adult informants. *Infant & Child Development: An International Journal of Research and Practice, 18*(1), 61-71.

Vélez, N., & Gweon, H. (2019). Integrating incomplete information with imperfect advice. Topics in *Cognitive Science, 11*(2), 299-315.

Weatherhead, D., White, K. S., & Friedman, O. (2016). Where are you from? Preschoolers infer background from accent. *Journal of Experimental Child Psychology, 143*, 171-178. https://doi.org/10.1016/j.jecp.2015.10.011

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: Bradford.

Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford University Press.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*(3), 655-684.

Wellman, H. M. & Gelman, S. A. (1992). Cognitive development: Foundational theories of core domains. *Annual Review Psychology, 43*, 337–375. doi:10.1146/annurev.ps.43.020192.002005

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*(2), 523-541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Wellman, H. M., & Liu, D. (2007). Causal reasoning as informed by the early development of explanations. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 261–279). Oxford, England: Oxford University Press. https://doi.org/10.1093/ acprof:oso/9780195176803.003.0017

Whitcombe, E. L., & Robinson, E. J. (2000). Children's decisions about what to believe and their ability to report the source of their belief. *Cognitive Development, 15*(3), 329-346.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103-128.