TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 00 (2026) e70067

ISSN: 1756-8765 online
DOI: 10.1111/tops.70067

This article is part of the topic "Developing Models of the World," Danaja Rutar, Johan Kwisthout, and Wanja Wiese (Topic Editors).

# An Introduction to Rational Constructivism in Cognitive Development

Rebekah A. Gelpi,[a] Pablo Leon-Villagra,[b] Daphna Buchsbaum,[b] Stephanie Denison[c]

[a]*Department of Psychology, University of Toronto*
[b]*Department of Cognitive and Psychological Sciences, Brown University*
[c]*Department of Psychology, University of Waterloo*

Received 31 January 2025; received in revised form 21 May 2026; accepted 22 May 2026

**Abstract**

Rational constructivism is a contemporary theory of cognitive development. It proposes that children generate, tweak, and radically revise theories, beliefs, and representations by integrating evidence with existing knowledge via probabilistic inferential learning mechanisms. We begin with a general overview of the theory of rational constructivism, covering its theoretical commitments and predictions. We then describe how it accounts for the empirically observed patterns of both incremental and radical developmental change and discuss the cognitive mechanisms underlying those conceptual changes. Finally, we sketch out a general computational framework for it and address open questions and directions for future research.

*Keywords:* Cognitive development; Rational constructivism; Child cognition; Computational cognitive development; Learning

Correspondence should be sent to Daphna Buchsbaum, Box 1821 Metcalf Research Building, 190 Thayer Street, Providence, RI 02912, USA. E-mail: daphna_buchsbaum@brown.edu Stephanie Denison, PAS Building 200 University Ave West, Waterloo, ON, N2L3G1, Canada. E-mail: smdeniso@uwaterloo.ca

## 1. Introduction

The fundamental question of cognitive development is how knowledge and cognitive capabilities change over the lifespan. How do we go from not knowing our own names and lacking concepts like *zero* and *more* to being able to use our knowledge to build bridges that stand for centuries or write clever song lyrics? Theories of cognitive development are broadly concerned with characterizing the starting state of the infant mind and exploring the mechanisms that support change over time. In this introductory paper to the special issue, we provide a general outline of rational constructivism—a theory of cognitive development. We situate it relative to other theories, describing its theoretical commitments and how it accounts for empirically observed patterns of developmental change. We then review the evidence for the mechanisms that support developmental change, detail a general computational framework for it, and highlight open questions and directions for future research.

Rational constructivism was developed as a theoretical framework for reconciling observed early emerging cognitive abilities in infancy with the profound cognitive change we see across childhood (see Xu, 2007, 2016, 2019; Xu & Griffiths, 2011; Xu & Kushnir, 2012). In doing so, it synthesizes three major theoretical approaches in cognitive development: constructivism, nativism, and empiricism. Constructivism, first developed by Jean Piaget (1954), drew from biological models of human development and argued that children actively construct the capacity for adultlike cognition through a gradual process of gathering new information and adjusting their representations and theories in turn. Piagetian constructivism postulated that children start from an initial state of only rudimentary sensorimotor representations in infancy, without any capacity for abstract, symbolic thought, and gradually build up to the rich, fully conceptual knowledge of adulthood. Accordingly, children must proceed through qualitatively distinct, stage-like changes in cognition as they overturn their earlier theories in favor of more well-supported hypotheses.

This relatively sparse characterization of infants' innate cognitive competencies is clearly at odds with the nativist argument, characterized by philosophers such as Plato, Renée Descartes, Gottfried Leibniz, and Immanuel Kant, that human beings are innately endowed with rich knowledge of concepts and capacities for reasoning (Laurence & Margolis, 2024; Samet, Bronstein, Brown, Helmig, & Normore, 2024). And empirically, Piaget's constructivist view has faced contemporary challenges inspired by such nativist philosophies, including large swaths of experimental data with infants and young children, showing that from a very young age, humans display extensive capacities for reasoning across a number of cognitive domains (for reviews, see Carey, 2009; Spelke, 2022). These findings of early cognitive capacities emerging in many cases years before Piaget had described, marked something of a "nativist turn" in cognitive development. Much empirical work supports a picture of human infants as possessing innate representationally rich structural and conceptual "core knowledge" (Carey & Spelke, 1994) that facilitates early reasoning in specific, evolutionarily important cognitive domains such as number, space, objects, and social agents. Further, this core knowledge has been shown to exist alongside the development of more complex cognitive capacities later in life, grounding the underlying ability to represent these domain-specific forms of knowledge throughout the lifespan.

Simultaneously, though, advances in the capabilities of artificial neural networks led to the development of an alternative empiricist account for children's early abilities. Like Piaget, these proposals argue that children are born with a lean starting state that includes very few built-in representations, but unlike Piaget, they propose a powerful, associationist learning mechanism. They argue that children's developing capacities throughout infancy and early childhood can be described by artificial neural networks sometimes referred to as connectionist models, which learn unstructured, non-symbolic latent representations of a given task (Elman et al., 1996; Munakata & McClelland, 2003).

While Piaget's specific claims were found to substantially underestimate infants' early abilities, recent work exploring many areas of core knowledge such as social cognition and number, as well as others such as language and meta-learning, has revealed substantial developmental change across infancy and early childhood. These instances of significant change in the nature of children's representations revived interest in *construction* as a motivating analogy for the process of cognitive development. In light of these findings, rational constructivist theories (e.g., Gopnik & Wellman, 2012; Xu, 2016, 2019; Xu & Kushnir, 2013) and related perspectives within probabilistic modeling of human cognition (e.g., Piantadosi, 2021; Rule, Tenenbaum, & Piantadosi, 2020; Ullman & Tenenbaum, 2020; Xu & Griffiths, 2011) explicitly articulate a set of proposed high-level mechanisms of change. On these theories, a child with a (minimal) set of innate knowledge develops working theories about the world by gathering information, testing their beliefs through active intervention, and revising or overturning their theories rationally via evaluating the strength of the evidence for and against different candidate theories. In doing so, they provide an explanatory model of belief revision and conceptual change that can be formalized more precisely than could Piaget's constructivist theory, which provided a fuzzier account of the specific mechanisms by which theory change occurs (e.g., Brainerd, 1978).

Although commitments to the specific nature of children's initial representational capacities and to the implementation of learning mechanisms may vary, we see three major commonalities uniting rational constructivist theories, broadly characterized:

1. People's concepts are theory-like representations that license predictions and explanations about the world, which must be testable and revisable. Our current set of theories constrain how we understand the world and interpret data, shaping our future theories.
2. Revision primarily occurs via statistical and inferential learning mechanisms that are *rational* in the sense that they approximate Bayesian inference, allowing for the updating of prior beliefs with new evidence and for drawing strong, inductive conclusions from observed data. Revision can take the form of either incremental or radical conceptual change.
3. Human beings possess the capacity and tendency to act on the environment to obtain and generate information efficiently and autonomously.

Throughout, we will use the term "rational constructivism" to describe this set of commitments. Next, we lay out the evidence supporting these claims, beginning by identifying discontinuities in infants' and young children's representations of everyday concepts, reflecting an overturning or revision of these earlier representations in favor of more flexible and generalizable knowledge.

## 2. Rational constructivism: Predictions and evidence

A key prediction of rational constructivism is that children should be capable of not simply tweaking but radically revising initially held beliefs, whether these initial beliefs take the form of conceptually rich core knowledge (e.g., Baillargeon, 2002; Carey, 2009; Spelke & Kinzler, 2007), innate theories (e.g., Gopnik & Wellman, 2012), or something straddling the line between perceptual and conceptual (e.g., "proto-conceptual primitives": Xu, 2016, 2019). Children show the capacity to reason about many domains, including numerical quantities, intuitive physics, and social agents, from early in infancy (Gopnik & Wellman, 2012; Spelke, 2022; Xu & Kushnir, 2013). However, these early-emerging capacities also differ substantially from those observed later in life, often exhibiting a discontinuous relationship as children acquire more complex, abstract reasoning abilities. Rational constructivism accounts for this phenomenon, observed across multiple domains of child cognition, by positing that children's changing behavior on these tasks across development reflects a conceptual revision and reorganization of children's naive theories.

For an example, let us consider empirical investigations of how infants represent kinds. Sortal kinds are entities that can be enumerated and individuated; "person" is a sortal kind since we can use information about the number of people, or about individuating characteristics of distinct people, to determine the identity of a person, while mass nouns such as "water" are not. Xu and Carey (1996) tested 10- and 12-month-old infants' abilities to individuate objects on the basis of kinds. During the experiment, one object (a duck) emerged from behind a central occluder and was then hidden. Then, another object (a ball) emerged from the occluder only to be hidden again. Finally, the occluder was removed, and children were allowed to see behind it. Because the identity of the objects presented differed in kind, a child tracking individuated objects should expect to see two objects when the occluder is taken away, even though they only ever saw one object at a time. Twelve-month-olds, but not 10-month-olds, looked reliably longer at the unexpected display (one object) than the expected display (two objects). One rational constructivist account of this failure is that younger infants' representations of hidden objects are less conceptually rich than older infants. Ten-month-olds might represent only that an object went out of view, but not the category membership of that object, while by 12 months infants may have a more enriched theory of objects that includes their kind. Critically, children whose parents indicated that they understood the meanings of the words used in the task (e.g., duck or ball) were more likely to look longer to the one-object display than the two-object display, suggesting a role for language acquisition as a potential driver of infants' development from a general concept OBJECT to a more mature object concept based on specific identities.

Another area in which children display very early understanding that is nevertheless subject to substantial change across development is biological reasoning. At its lowest level, reasoning about biological concepts requires distinguishing between entities that govern their own actions and motions and those that do not. One of the earliest and most relevant distinctions that infants can use to reliably establish agency is between biological and nonbiological motion. Even newborns in the first few days of life recognize these forms of motion as distinct (e.g., Bardi, Regolin, & Simion, 2011; Simion, Leo, Turati, Valenza, & Barba, 2007); this capacity is shared with visually naive newborn chicks (e.g., Vallortigara, Regolin, & Marconato, 2005) and other species including cats and fish (Blake, 1993; Schluessel, Kortekamp, Cortes, Klein, & Bleckmann, 2015), and by the end of the first three months of life, infants consistently prefer biological motion displays to visually similar nonbiological motion displays (Bertenthal, Proffitt, & Cutting, 1984).

While early categorization based on cues like motion gets infants off the ground in biological reasoning, these basic notions are revised and reorganized into more theoretically rich and complex structures as evidence accumulates to suggest this early organization is too simplistic. Young children begin to use their knowledge of intuitive biology to understand what unites living things, such as growth, motion, and illness (e.g., Carey, 1985; Inagaki & Hatano, 2004; Wellman & Gelman, 1992). For example, starting around age 3 years, children infer that members of a biological category share an internal essence that causes them to display category-typical properties and behaviors, even when they do not appear superficially similar to other members (Gelman & Davidson, 2013; Gelman & Markman, 1986; Gelman & Wellman, 1991). They also begin to shift from a view that biological categories are mutually exclusive to one another (Markman, 1989) to a hierarchically structured taxonomic theory of biology, using superordinate categories such as "plant" or "animal" (e.g., Waxman & Gelman, 1986) as well as subordinate categories such as "oak tree" appropriately based on context such as contrasting information (Waxman, Lynch, Casey, & Baer, 1997; Xu & Tenenbaum, 2007) and linguistic form (Gelman, Wilcox, & Clark, 1989). These revisions to biological theory represent significant conceptual change, where children restructure the contents of categories, as well as category structure itself when evidence in the world contradicts their earlier theories.

Relatedly, children also exhibit substantial developmental change in their reasoning about social agents. Infants use cues such as animacy to reason about the agency of objects in their environment (e.g., Schlottmann & Ray, 2010), developing an understanding that the actions of such agents are goal-directed (Csibra, Bíró, Koós, & Gergely, 2003; Gergely & Csibra, 2003; Woodward & Sommerville, 2000; Woodward, Sommerville, & Guajardo, 2001). These cues facilitate an early understanding of others as social agents who act intentionally on the world. Further, although other human beings are an understandable focus of children's early life, cues to agency can be applied widely, as infants and children attribute agency to inanimate objects such as puppets, at least when treated as agents by others (Asaba, Li, Yow, & Gweon, 2022; Yu & Wellman, 2022, but see Dragon & Poulin-Dubois, 2023). But again, this early understanding is in some ways limited. Empirical work suggests that infants' initial theory of desires assumes that preferences are universal across people (or at very least, that other people's preferences will align with the infant's). For example, on a task in which an adult expresses a dislike for a food the child likes, infants tend to behave as though the adult's preference is indeed aligned with theirs, offering the adult food that they themselves like, but the adult professed a dislike for (Repacholi & Gopnik, 1997). But as children develop, they are faced with increasingly varied situations in which they observe individuals whose goals and intentions sometimes align with, but sometimes differ from, the child's own. Thus, by the ages of 2 to 3 years old, toddlers exhibit an understanding that others can hold preferences distinct from their own (Doan, Friedman, & Denison, 2021; Ma & Xu, 2011). And importantly, infants who are exposed to additional experiences with differing preferences through a training procedure show this more differentiated theory of preferences in their behaviors (Doan, Denison, Lucas, & Gopnik, 2015). Together, these findings suggest that young infants go from a parsimonious hypothesis—that is, that preferences are universal or identical to the infants' own—but are capable of revising this hypothesis as they encounter additional evidence that preferences, although often similar, are not universally shared (Lucas et al., 2014).

Related to this broadening understanding that the underlying causes of human behavior are often abstract intentions, goals, beliefs, and desires, children also learn to represent abstract physical causal relations and use this understanding to intervene on their environments (for reviews, see Goddu & Gopnik, 2024; Muentener & Bonawitz, 2017). Children's causal reasoning abilities develop significantly in the first few years of life: As infants and toddlers they distinguish causal relations from non-causal ones (Cohen & Amsel, 1998; Leslie & Keeble, 1987; Saxe & Carey, 2006; Tecwyn, Mazumder, & Buchsbaum, 2023) and use probabilistic inference to reason about causes (Gopnik & Schulz, 2004; Sobel & Kirkham, 2007). In the toddler and preschool years, they observe and then imitate the intentional actions of others to achieve causal outcomes and more generally develop their capacity for engaging in causal interventions (Buchsbaum, Gopnik, Griffiths, & Shafto, 2011; Gergely, Bekkering, & Király, 2002; Meltzoff, 1995; Schwier, van Maanen, Carpenter, & Tomasello, 2006). Indeed, by the time children are roughly 5 years old, they have gained ample knowledge about how their actions affect the world through observation, instruction, exploration and play (Schulz, 2012). Accordingly, much literature shows that children design interventions on causal systems and learn from them to reveal relationships between causal variables (Bramley, Jones, Gureckis, & Ruggeri, 2022; Kushnir & Gopnik, 2005; Lapidow & Walker, 2020; Ruggeri, Swaboda, Sim, & Gopnik, 2019; Schulz et al., 2007; Sobel & Sommerville, 2010). But what mechanisms allow children to revise their theories in these domains (e.g., psychological and physical causality) so dramatically?

### 2.1. *Mechanisms of theory change across development*

One of the ways in which rational constructivism distinguishes itself from some previous constructivist theories is in its specification of mechanisms and systems that facilitate knowledge acquisition and change. Specifically, rational constructivism assumes that humans are by and large *rational* learners, with respect to their current set of beliefs (often referred to as hypotheses). This assumption of rationality postulates that human beings integrate new evidence with their current belief states proportional to the strength of their candidate beliefs and the likelihood of the observed evidence under those beliefs, a computation formally equivalent to Bayes' rule (for an introduction, see Perfors et al., 2011).

To give an example, someone flipping a coin might assume that the chance of encountering a trick coin that always lands on heads rather than a fair coin is very low in general; say, 1 in 1 million. But let us say you flip a coin 10 times and each time it lands on heads. This is more likely with a trick than a fair coin, and so you would increase your belief that it is a trick coin accordingly. However, because the prior probability of it being a trick coin is so low, you should still believe that the coin is most likely fair, say upping the chance of it being a trick coin to 1 in 1000. However, if the coin lands on heads 100 times in a row, you would now be pretty convinced that something fishy is happening because the chance of that happening with a fair coin is vanishingly rare.

However, as Bayesian reasoning is often computationally intractable in real-world settings with many candidate hypotheses and a large amount of information to integrate, both children and adults are often not able to engage in idealized Bayesian reasoning (Bhui, Lai, & Gershman, 2021; Lieder & Griffiths, 2020). Nevertheless, Bayesian models of cognition have predicted human performance on a variety of tasks (e.g., Austerweil, Gershman, Tenenbaum, & Griffiths, 2015; Goodman, Ullman, & Tenenbaum, 2011; Ullman & Tenenbaum, 2020), and computational cognitive science has extended these theories by proposing models with more biologically plausible assumptions about cognitive resource constraints (e.g., Bramley, Dayan, Griffiths, & Lagnado, 2017; Griffiths, Chater, Kemp, Perfors, & Tenenbaum, 2010; Lieder & Griffiths, 2020). In this section, we review several mechanisms that facilitate children's capacity to revise their beliefs across development: probabilistic inference, inductive constraints, abstraction, and hypothesis generation.

Given the centrality of probabilistic reasoning to Bayesian models of cognition, rational constructivism predicts that from infancy, humans should be able to detect and evaluate probabilistic information in the environment and use these probabilistic patterns to make inferences and predictions. Converging evidence from multiple domains suggests that infants can use statistical information within the first year of life to reason about their environment. For example, infants can use the transitional probabilities of syllables to detect word boundaries (Saffran, Aslin, & Newport, 1996) and visual patterns (Kirkham, Slemmer, & Johnson, 2002), providing them with a critical early tool for segmenting regularities in language and the visual stream. Infants can also make inferences about probabilistic outcomes, looking longer at improbable sampling outcomes, and toward the end of the first year, using probability to guide choices (Denison & Xu, 2010; Denison, Reed, & Xu, 2013; Teglas et al., 2007; Xu & Garcia, 2008).

Critically, infants between 1 and 2 years old are also developing a sense of how seeming "violations of probability" are not simply unusual surprises but can reflect differences in the data generation process. This manifests as an early curiosity by 13-month-olds to preferentially explore samples that seem "suspiciously" discordant with the underlying probability based on the distribution from which they are drawn (Sim & Xu, 2017). Infants also begin to reason about the explanation for these non-random samples: For example, if a jar contained mostly blocks instead of spring toys, a child might infer that someone who pulled only spring toys out of the jar for themselves was not simply randomly drawing the rarer item but was specifically intending to take out the spring toys, possibly due to the person's preference (e.g., Gweon, Tenenbaum, & Schulz, 2010; Kushnir, Xu, & Wellman, 2010). This early understanding that intentionally generated data exhibit different patterns than randomly generated data, in turn, facilitate children's ability to make stronger inferences about how other people's preferences can differ from their own (Lucas et al., 2014; Ma & Xu, 2011) and how (or how not) to learn from teaching (e.g., Bonawitz et al., 2011; Buchsbaum et al., 2011; Shafto, Goodman, & Griffiths, 2014).

More broadly, children's learning from social partners does not only take into account the perceived intentions and desires of the demonstrator but also inferences that children make about the demonstrators themselves, such as their reliability, informativeness, and confidence (Bridgers, Buchsbaum, Seiver, Griffiths, & Gopnik, 2016; Buchsbaum, Bridgers, et al., 2012; Buchsbaum, Seiver, Bridgers, & Gopnik, 2012; Gelpí & Buchsbaum, 2025). By the time children are 3 years old, they are beginning to develop a sense of selective trust, endorsing testimony more frequently when an informant is known to be an expert in the domain in question or has previously been accurate, as well as when an option is endorsed by a consensus or a majority of (independent) informants (for reviews, see Gelpí & Buchsbaum, 2025; Harris, Koenig, Corriveau, & Jaswal, 2018; Mills, 2013). Children also tune their inferences according to the domain in which they are learning, showing a greater sensitivity to consensus in normative domains than causal domains (Pham & Buchsbaum, 2020), and extending more credibility to testimony that conflicts with their own ambiguous observations when an informant is confident rather than naive (Bridgers et al., 2016), although the strength of this tendency varies across cultural contexts (Sebastián-Enesco et al., 2020). They consider not only the number of endorsements but the source of their informants' and their own knowledge, though they do not always weight these sources as an adult would (Gelpí, Otsubo, Whalen, & Buchsbaum, 2025; Gelpí, Whalen, Griffiths, Xu, & Buchsbaum, 2025). These cues allow children to further narrow their hypothesis space by evaluating the likelihood of information within its social context, discarding social information that is perceived to be less reliable, and placing greater weight on social information from more trusted sources, particularly in domains where asocial learning is impossible.

Another cognitive mechanism that complements probabilistic reasoning and social learning is children's ability to identify abstraction. Developing abstract, hierarchical concepts for the relationships that (learned) statistics map to can greatly simplify the representational complexity of the environment. For example, a thought experiment from Goodman (1955) presents a situation in which one has a number of bags, which have marbles of unknown color inside. Opening up one bag and removing all the marbles reveals that every marble in this first bag is red. Opening up the second bag reveals that every marble in this bag is blue. After drawing a single marble from the third bag, and seeing that it is yellow, a plausible hypothesis is that the remaining marbles in this bag are yellow, even though most of the other marbles observed thus far have been red or blue. Further, a learner might infer that all unopened bags, while having unknown colors of marbles inside them, may be more likely to each have the same color of marble inside them. These hypotheses about hypotheses, or *overhypotheses*, can in principle be learned quickly and applied to new situations, making children and adults alike more effective probabilistic learners (Dewar & Xu, 2010; Felsche, Stevens, Völter, Buchsbaum, & Seed, 2023; Felsche, Völter, Herrmann, Seed, & Buchsbaum, 2024; Goodman et al., 2011; Kemp, Perfors, & Tenenbaum, 2007).

Learning abstract symbol systems, most prominently language, may also undergird children's generalization about other systems. One prominent example can be seen in the development of number concepts: Children who learn languages with exact number systems undergo a transition from learning to count to 4 or 5 to a general understanding that all integers represent a unique, exact quantity (Carey, 2009; Piantadosi, Jara-Ettinger, & Gibson, 2014), while speakers of languages without larger number concepts shift to using approximate number systems for larger quantities (Pica, Lemer, Izard, & Dehaene, 2004). Language may also play a key role by allowing children to generalize in sophisticated ways about early-emerging, domain-specific knowledge. When children learn about biological categories, for example, they are often presented with information like "birds fly" or "mammals have fur" (Gelman & Raman, 2003; Gelman, Goetz, Sarnecka, & Flukes, 2008). These generic statements license broad generalizations about category members (Cimpian & Markman, 2009; Gelman, Ware, & Kleinberg, 2010), even when a category member is atypical in other ways (Graham, Gelman, & Clarke, 2016). Importantly, however, these statements are flexible, such that category members can also sometimes violate a general pattern without the generic statement being wrong—for example, that it can be true that birds can fly and that penguins are birds, despite penguins not being able to fly (Gelman & Raman, 2003). Children use and understand generic statements as young as age 3 (Gelman et al., 2008; Gelman, Leslie, Was, & Koch, 2015), and these statements provide children with a powerful inductive tool to organize their world into natural categories, or "kinds" (Brandone, Cimpian, Leslie, & Gelman, 2012; Khemlani, Leslie, & Glucksberg, 2012). Importantly, the salience and importance of a category depends on one's social environment; thus, children's induction must be (and is) flexible enough to allow them to learn culture- and language-specific category knowledge (see also Diesendruck, 2003).

All of the above mechanisms provide children with a rich toolbox for revising their existing beliefs given the data they encounter and inductive principles that allow for broad generalization and rapid learning across all domains. However, in addition to requiring mechanisms that allow them to induce the likeliest hypotheses out of a set of candidates after observing data, children must be able to generate new hypotheses to evaluate in the first place. One such tool that could enable this is a capacity for "constructive thinking" (Lombrozo, 2020; Xu, 2019), allowing for not only incremental revision of existing beliefs but the generation of ideas outside of one's current hypothesis space. As relative newcomers to the world, children lack a great deal of explicit knowledge and must learn to evaluate what they do not know and how to obtain this information. Sometimes, this can take the form of thinking through explanations for the data they encounter in the world; for example, 5-year-old children prompted to provide an explanation for an outcome (whether a garden is healthy or sick) were more likely to generate parsimonious hypotheses for the data than those prompted to report the outcome (Walker, Bonawitz, & Lombrozo, 2017), and investigations of adults have similarly revealed that the explanation drives their hypothesis generation (Brockbank & Walker, 2022). These efforts to seek explanations for the data they observe also drive children's active information search through other means, such as asking questions. And finally, children can often take advantage of evidence in the form of explanations from others (e.g., Ganea, Larsen, & Venkadasalam, 2021), quickly revising their beliefs when these explanations prove to be more effective and robust than their initial theories.

Additionally, these mechanisms—probabilistic inference, inductive biases, abstraction, and hypothesis generation—work in concert to allow children to assess evidence and revise their beliefs in the face of data (Bonawitz, Ullman, Bridgers, Gopnik, & Tenenbaum, 2019). In one study, children were presented with a task in which they must attempt to reason about the rules governing objects with a novel causal property similar to magnetism (blue or yellow objects, in which like colors repel one another and opposites attract) or "inverted" magnetism (i.e., like colors attract, and opposites repel). Children were initially shown ambiguous evidence: three blocks of unknown color stuck to blue blocks and repelled yellow blocks, and three unknown blocks stuck to yellow blocks and repelled blue blocks. Because the identity of these blocks was not known, the data at this point were consistent with both the magnetism theory and the inverse-magnetism theory. Then, they were presented with a single piece of disambiguating evidence that was either consistent with magnetism (e.g., a blue block sticking to a yellow block but pushing away other blue blocks) or the inverse-magnetism theory (e.g., two blue blocks sticking together but pushing away yellow blocks). Despite having a strong prior belief toward hypotheses favoring "sticking" relationships (i.e., that all blocks will stick), children successfully generated the two possible candidate hypotheses significantly more often than any others and used the disambiguating evidence to come to the correct solution based on the disambiguating data more often than other potential solutions. This ability to revise beliefs to account for new evidence can occur even when the theories supported by the evidence seem counterintuitive to young children, such as germ theory or psychosomatic illnesses (Keil, Levin, Richman, & Gutheil, 1999; Schulz et al., 2008).

### 2.2. *Active learning*

The idea that children actively seek evidence from their environment to construct and revise their intuitive theories, and to develop effective tests of these theories, often termed *active learning*, plays a central role in rational constructivism. In this framework, the active nature of children's learning reflects not just learning from motor behavior and feedback but the way in which children actively observe the world around them, whether by asking questions, selecting information to learn from, or intervening on causal systems.

From early infancy, even before developing the capacity for fine motor control, humans exhibit selective attention, preferentially attending to some stimuli over others and exploring them more deeply and using the statistical contingencies of this information to form rudimentary categories of sameness and difference (Oakes, Madole, & Cohen, 1991). Additionally, infants are drawn to visual and auditory content that sits in a "Goldilocks" zone of medium complexity. This pattern of attention allows them to update their beliefs to accommodate new information by focusing on events that are moderately surprising and avoiding levels of probability that are so low there is little to be learned or so high that integrating it with existing information is difficult (Kidd, Piantadosi, & Aslin, 2012; Kidd, Piantadosi, & Aslin, 2014).

Infants also appear to evaluate data with multiple candidate hypotheses in mind, excluding a theory only when data contradict it (Gerken, 2010) and balancing their inferences about the likelihood of the data given a hypothesis with the parsimony of the hypothesis (Bonawitz & Lombrozo, 2012). Given that children are still learning a great deal about the world, and many of their theories must continue to be developed and revised throughout childhood, a tendency for "creative rationality" (Fedyk & Xu, 2018; see also Giron et al., 2023; Gopnik, 2020; Pelz & Kidd, 2020), in which children might test very unusual, low-probability, or counterfactual hypotheses despite their seeming implausibility or irrelevance to the task at hand, might also be a useful way for them to gain knowledge about the world. Although this principle conflicts in some ways with principles of idealized Bayesian rationality, it may nevertheless optimize long-run information gain, as children must often radically revise their theories in order to learn potentially counterintuitive, but correct, explanations for phenomena such as density or gravity.

Although children learn actively from infancy, their emerging capacities for coordinated action increasingly enable them to explore their environment and change it through their own interventions. A key dilemma that young children face is the information selection problem: Knowing what they do and do not know, and which actions will result in outcomes that fill in the gaps in their knowledge. Particularly when they have little task-specific knowledge to guide their choices, young children initially explore their potential options—sometimes they do so randomly (Giron et al., 2023; Meder, Wu, Schulz, & Ruggeri, 2021; but see Schulz, Wu, Ruggeri, & Meder, 2019), sometimes systematically (Blanco & Sloutsky, 2021; Sumner, Steyvers, & Sarnecka, 2019), but in general, they appear to spend more time learning about their environment than exploiting well-known areas of reward.

Although both random and systematic exploration can provide ample information, these strategies often do not yield optimal outcomes when the opportunities to learn are particularly sparse or when obtaining rewards is more important. In these situations, directed exploration toward areas known to be informationally rich or highly uncertain is an important tool for improving rapid learning and long-run reward. As mentioned earlier, children's and adolescents' ability to design domain-specific interventions that optimize their learning improves over time (e.g., Nussenbaum et al., 2019; Schauble, 1990). Similarly, children learn strategies that help them resolve uncertainty in pure-exploration environments (e.g., Ruggeri, Stanciu, Pelz, Gopnik, & Schulz, 2024) and environments that trade off exploration and exploitation (Meder et al., 2021; Schulz et al., 2019).

One of the most complex domains where children must develop strategies for obtaining information is natural language. Unlike many experiments or games, in which children's action space is highly limited, there are a theoretically infinite number of questions that children can ask, making it particularly important for children to identify the features of questions that result in receiving relevant and informative explanations. Although younger children struggle to both generate informative questions and recognize when questions no longer provide additional information, even 5-year-olds can identify informative questions when choosing from options provided to them (Ruggeri & Lombrozo, 2015; Ruggeri, Sim, & Xu, 2017). Also, children's overall ability to generate and learn from question-asking improves with age (Jones, Swaboda, & Ruggeri, 2020; Ruggeri & Lombrozo, 2015), possibly because they develop strategies that can be reused or tweaked across different situations (Liquin, Rhodes, & Gureckis, 2024). We will return to how children might implement such a strategy below, but first, we turn our focus to specific algorithms of theory change.

## 3. Algorithms of theory change

So far, we have presented findings of developmental theory change and discussed how these findings are characterized at a high level by the mechanisms proposed by rational constructivism. In this section, we will showcase how recent computational approaches can capture these empirical results and be used to critically assess the theoretical commitments of rational constructivism. As we have established, learning and development start with a set of building blocks and an agent's ability to construct naive theories of the world from these initial elements through inference and learning. These intuitive theories go beyond perceptual stimuli, positing invisible latent variables and abstract principles.

As discussed above, learning in the rational constructivist framework operates at multiple hierarchical levels of abstraction (see Goodman's, 1955, thought experiment). At the theory level, the agent is learning what theories and representational structures are relevant for different domains: For example, different species of animals can be organized along taxonomic structures, whereas social relationships can be captured by the connections of social network graphs. At a more concrete level, details of the structures of these theories must also be learned: A child must learn that, while they are perceptually similar, sharks and dolphins should be positioned at different branches of the taxonomy.

In the rational constructivist framework, learning both at the theoretical and concrete level amounts to Bayesian updating: using observations combined with our prior beliefs to derive a new belief via Bayes' rule. This very general framework aligns with Marr's (1982) computational level of analysis, characterizing the problem of how agents ought to update their beliefs when faced with new data and providing an optimal solution to the task. Comparing how closely humans match these ideals or describing how their inferences deviate from the optimal solutions has been a very fruitful perspective in cognitive science (e.g., Oaksford & Chater, 2007; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). However, computational-level theories do not capture the processes and algorithms by which a solution is reached (the algorithmic level) or the method by which these processes are physically realized (the implementational level).

We have seen that the naive theories children have allow for flexible inferences and a range of behaviors. For example, the theories children learn about animals can inform flexible generalizations about which features all members of an animal group, such as birds or fish, should share. They should also allow children to reason about novel, and often latent, properties these animals could exhibit, and classify novel animals as belonging to the same category or not (Kemp & Tenenbaum, 2008, 2009). Moreover, these behaviors are not limited to feature or category inferences and generalizations; children use these theories to ask questions (Liquin et al., 2024), predict how actions or interventions would change an entity or category member (Kemp et al., 2007), or simulate behaviors or events through play (Chu & Schulz, 2020).

The range of behaviors that these theories can facilitate suggests that the theories that children learn are not simply rules that discriminate between, for example, category members like birds (and non-birds), but full-fledged models of the domain in question that license predictions, novel inferences, explanations and generalizations (see Fig. 1). These *generative models* attempt to approximate how objects, agents, behaviors, or events in a domain are

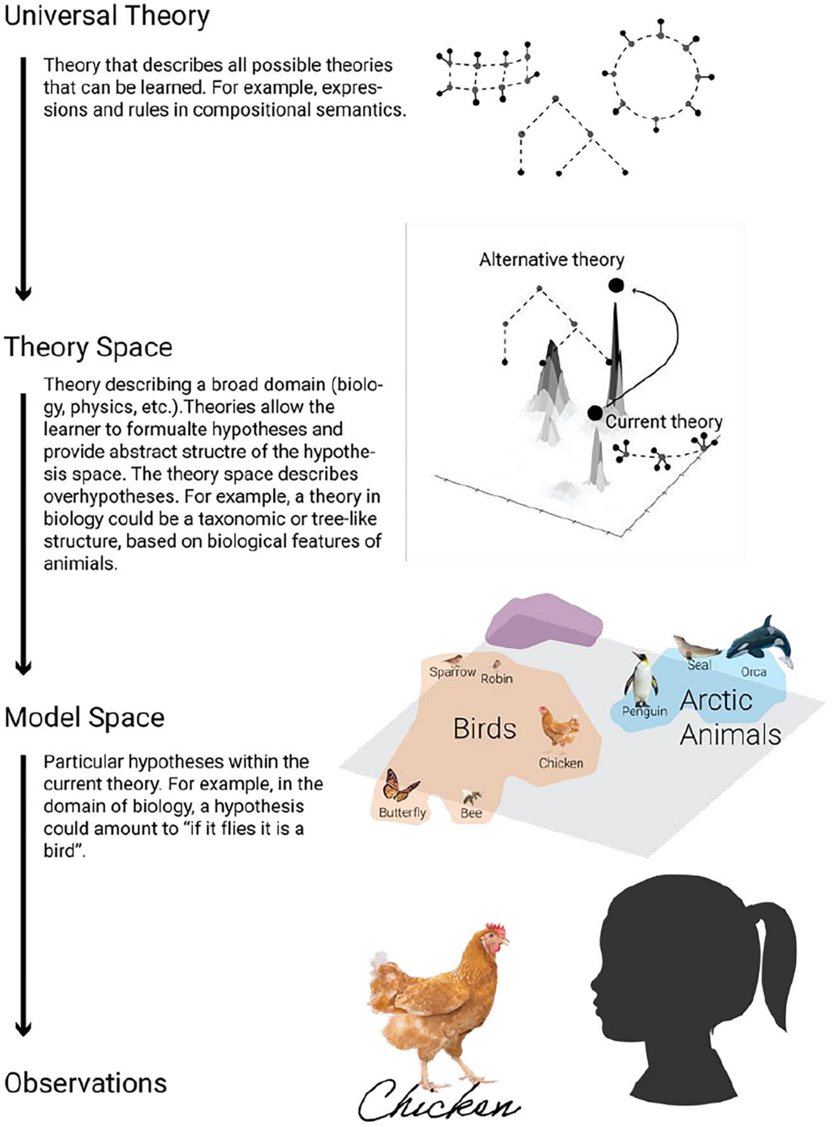


Fig. 1. An example of a hierarchical theory space. Theories at each level influence the structure and probability of hypotheses at the level below, and observations are in turn used to update the theory spaces.

produced. For example, when describing how objects move in their environment, a naive theory of physics would attempt to model the forces governing the movement of objects, such as by inferring magnetic forces when observing how magnets can displace other objects (e.g., Bonawitz et al., 2019) or how properties such as mass, friction, and gravity jointly interact within complex scenes (Ullman, Stuhlmüller, Goodman, & Tenenbaum, 2018). Similarly, a naive theory of biology could posit latent factors that generate the different appearances or behaviors of different kinds of animals and that some, but not all, externally observable characteristics such as shared features or behaviors reflect taxonomic proximity.

These generative theories are flexible since they can make predictions about novel entities (e.g., inferring whether a flying animal you have never seen before is a bird) and produce new data by generating how different assumptions would result in, for example, different physical movements or different biological organisms. This flexibility allows a learner's theory to support a range of behaviors, from inference and generalization to question asking or planning of causal interventions. However, this flexibility also comes at a price— learning these theories is data intensive and computationally challenging.

Computational models of rational constructivism must provide accounts of how these different representations can be learned from simpler building blocks and how these representations allow the learner to infer and generalize. As we touch on in our discussion of innate knowledge, multiple sources of initial knowledge have been suggested, such as core knowledge, but so far there is no consensus on what exactly these initial concepts consist of or how expressive and adaptable they are. However, to allow the agent to construct theories from these initial elements that can be used to derive inferences, these elements must be composable, that is, the agent must be capable of constructing more complex theories from the initial concepts. This form of composing more complex expressions from simpler elements resembles how languages can express propositions about the world by composition, and one well-known theoretical and computational perspective on the nature of compositional representations is that these naive theories are produced by a language of thought (LOT, Fodor, 1975),[1] though compositionality may be implemented in other ways.

In the following subsections, we will first discuss different formalisms that can function as structured knowledge representations, and how these allow for theory-like knowledge representations. Then we will show that learning these knowledge representations is a challenging computational task and discuss potential mechanisms to overcome these challenges.

### *3.1. Learning structured knowledge representations*

We have seen that different types of structured representations have been suggested to account for the richness of human concepts and intuitive theories. One central issue that computational models of rational constructivism must address is how these structures can be learned, particularly given the theoretically unbounded nature of the hypothesis space over all possible structures.

Early computational accounts of children's learning tended to sidestep the question of how theories could be learned, instead focusing on how children derived inferences using existing theories and how their inferences compared to the Bayesian ideal (e.g., Griffiths, Sobel, Tenenbaum, & Gopnik, 2011). More recent extensions approach theory learning but simplify it as deciding between sets of potential hypotheses or learning how the observed data could be explained by introducing latent, unobserved features (Kemp & Tenenbaum, 2008, 2009). While these models have begun to describe how children can decide between alternative theories, and emphasized the role of abstract, hierarchical knowledge, these accounts still fail to capture how these candidate theories can be learned in the first place.

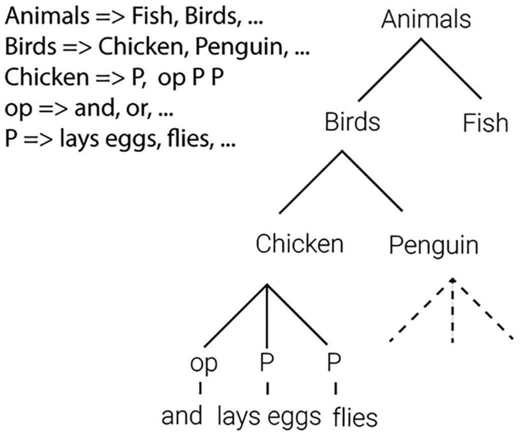


Fig. 2. An example of a context-free grammar for animal concepts.

Given the similarity of structured knowledge representations and language, one popular account of how theories could be constructed is inspired by computational formalisms of syntactic structures in language and assumes that learning theories can be formalized as learning a grammar over the probability of syntactic rules: in this case, a syntax describing the theoretical domain. Many such approaches use *probabilistic context-free grammars* (PCFGs) to construct complex knowledge representations (see Fig. 2; e.g., Bramley & Xu, 2023; Gelpi et al., 2020; Goodman, Tenenbaum, Feldman, & Griffiths, 2008). These PCFGs consist of grammatical rules that describe how symbols can be combined or expanded to produce more complex expressions and the probability of the new expression. For example, learning a rule such as "to activate the machine, the block has to have red stripes" could be derived from a set of rules such as "S $\rightarrow$ A and B, S$\rightarrow$ A or B, A $\rightarrow$ is red, is green, is blue, B $\rightarrow$ has stripes, has dots, and so forth." After seeing multiple red and striped blocks activating a machine, the probability of S$\rightarrow$ A and B; A $\rightarrow$ is red, and B $\rightarrow$ has stripes would increase (e.g., Bonawitz, van Schijndel, Friel, & Schulz, 2012; Goodman et al., 2008). Note, however, that in these settings, all the rules are fixed. Similar to computational models of language, more flexible models have been proposed that allow learning new rules (e.g., adaptor grammars, Zhao, Bramley, & Lucas, 2022; Zhao, Lucas, & Bramley, 2024).

While the building blocks of these PCFGs can be fairly simple, with the appropriate set of grammatical rules, an infinite set of structures can be generated—making infinite use of finite means. Note also that this formalism is not limited to one type of structure—given the right rule set, structures like trees, graphs, maps or even novel structures can be produced (Kemp & Tenenbaum, 2009; Lake, Lawrence, & Tenenbaum, 2018); similarly, such grammar-based models can capture learning of domains beyond categorization and concept representation. For example, program synthesis methods (e.g., Lake et al., 2015; Ellis et al., 2021) learn a library of composable strategies that can be reused to solve a variety of problems.

Learning in these setups has a hierarchical structure; one level of learning includes learning the right structure (graph, tree) for the problem and the second learning the particular structure (a particular graph, a particular tree) and its parameters. Relatedly, because a single syntactic rule change at the structure learning level can in turn restructure whole sub-trees of hypotheses, hierarchical grammars can model theory change—such as the shift from flat partitions of animal categories (Fig. 1) to a more taxonomical structure (Kemp & Tenenbaum, 2008).

A related formalism that can express complex knowledge from simpler constituents, and that allows for inferences, is propositional logic. Initially termed "laws of thought" (Boole, 1854), logic produces statements by combining propositions with logical connectives such as conjunctions or disjunctions. Crucially, expressing knowledge as logical propositions allows the learner to derive inferences, such as inferring that a red block will activate a music box from "if the block is red, it is a blicket" and "blickets activate the music box." To be able to derive these types of inferences, an agent would have to learn statements such as "if the block is red, it is a blicket" from simpler constituents such as "block" or "blicket" (which themselves might be learned from simpler constituents such as "a block is a solid and rectangular object") and (possibly innate) connectives such as "and" or "if…, then." In this view, rational constructivism amounts to learning structured knowledge representations from data, cast in propositional logic.

While propositional logic allows us to express complex relationships, it is still too limited to capture the range of representations humans can learn since propositional statements are either true or false; for example, the block is either a blicket or it is not, the box is red, or it is not (although other logic formalisms exist that can capture graded relationships, such as fuzzy logic). Still, in reality, these statements are often graded ("this block is likely a blicket," "I am not sure if the block was red"; see also Tessler & Goodman, 2019). Thus, many computational models compatible with the rational constructivist framework include ways of expressing graded knowledge (e.g., "most birds fly"). These can be most straightforwardly expressed in probability theory as distributions over possible values, and importantly these distributions can be combined and composed to produce more complex knowledge representations. One way to intuitively represent these complex relationships between probability distributions is in probabilistic graphs. Finally, when these graphs represent causal relationships, these probabilistic graphs can be augmented by including the direction of relationships between variables (A causes B) and specify how interventions affect the relationships in the graph (Pearl, 1995).

Critically, children also learn to use explicit symbolic systems that may facilitate the acquisition of concepts that are complex in meaning, polysemous, or otherwise difficult to represent. For example, children begin to use words conveying modal concepts (e.g., "could," "must") and logical disjunction (e.g., "or") in adult-like ways before they exhibit adult-like patterns of comprehension (Grigoroglou & Ganea, 2022). Importantly, these terms can carry multiple possible meanings or interpretations, and children tend to master simpler, less abstract meanings of these words first. Polysemous terms in natural language could "bootstrap" children's development of mature representations for concepts (e.g., Carey, 2009) such as possibility and disjunction by encouraging the generalization of a simpler concept to a broader set of circumstances.

We have seen that these computational models learn at least two levels: the abstract level of learning different knowledge structures, and the concrete level that describes the details of a particular structure. One important insight is that learning in these hierarchical structures can sometimes produce faster learning at the more abstract rather than the more concrete level (the "blessing of abstraction," Gershman, 2017; Goodman et al., 2011). As mentioned previously, these abstract hypotheses about how concepts tend to be structured, or overhypotheses, may be a particularly important element of children's ability to rapidly draw inferences about their environment after just a few observations. Kemp et al. (2007) presented a computational model that proposed how human beings can develop these overhypotheses by reasoning about the appropriate degree to which they should generalize based on the data they encounter. For example, as in Goodman's (1955) thought experiment discussed above, encountering multiple bags of marbles, each of which contains marbles of all one color, should increase an individual's confidence that (1) they can infer the color of the remaining marbles in a bag after seeing a single marble from that bag, and (2) they can infer that bags of marbles tend to all have one color, even if the specific color is unknown.

Infants' looking behavior suggests that they form similar overhypotheses and are surprised if a new container has mixed contents when containers in general are uniform (Dewar & Xu, 2010). Recently, Felsche et al. (2023, 2024) also showed that while preschool children and, to a lesser extent, chimpanzees were able to both learn and act on overhypotheses, capuchin monkeys could not, suggesting that this type of abstract learning might be a more recently evolved capacity.

One of the most important empirical examples is the so-called *shape bias*, the abstract belief that object categories are often defined by the shape of the category members instead of other features such as size or color (Landau, Smith, & Jones, 1988; but see Imai & Gentner, 1997; Jara-Ettinger, Levy, Sakel, Huanca, & Gibson, 2022; Lucy & Gaskins, 2001, for examples of cultural variation in the strength of the shape bias). This bias has also been found in artificial neural networks trained on object categories, suggesting that shape features are correlated with object categories (e.g., Ritter, Barrett, Santoro, & Botvinick, 2017). Inductive biases about the abstract structure of category concepts—such as that solid objects tend to be categorized by their shape rather than their color or other perceptual features—can considerably improve inference about a novel object or category since they narrow the learner's hypothesis space to focus on more promising theories of categorization. For instance, the learner does not have to consider the much larger set of features that define a novel object category, instead focusing on a feature (shape) that is generally highly predictive of categories for their given language or cultural context.

While the hierarchical structure of the learner's hypothesis space can already produce favorable learning dynamics, such as the emergence of overhypotheses, in general, the computational task of learning these representations is still extremely challenging. To illustrate the challenge, consider a child trying to learn what objects can be displaced by magnets. Learning the right knowledge structures consists of learning that the domain of physics can be described by laws that assume invisible forces that determine how objects in our environment interact. These objects have multiple features, such as their shape, color, or mass, as well as invisible properties such as "is magnetic." The task of the agent then is to infer the best hypothesis describing their observations, for example, "magnets displace round and small and shiny objects" or "magnets displace round and shiny objects" or "magnets displace metal objects." Importantly, there are many potential properties and laws describing objects at different degrees of specificity (e.g., do magnets only displace "round and shiny and red and light and…" objects).

The example illustrates that the hypothesis space contains an infinite space of properties (e.g., infinite conjunctions of properties). Moreover, finding the best solution in this space is extremely challenging since there is no guarantee that there will be other hypotheses that, while not completely correct, are "close" in structure to the correct solution, that is, the "knowledge landscape" may consist of a few disjointed plausible hypotheses (Ullman, Goodman, & Tenenbaum, 2012). Finally, learning hierarchical knowledge is challenging since the learner requires both the right abstract knowledge representation (a theory-like representation) and details of this representation (a particular hypothesis that explains an observation). But both theory and hypotheses interact: without the right hypothesis, no good theory can be learned, and vice versa—a "chicken and egg problem" (Bonawitz et al., 2019; Ullman et al., 2012).

While finding good hypotheses in these inference problems is an important open problem in machine learning, several promising approaches have been tested. One approach is based on tempering or annealing the "knowledge landscape"; that is, artificially reducing the difference between hypotheses with low explanatory power and high explanatory power. The idea, motivated by an analogy to metallurgy, is that searching for the highest point in this knowledge landscape (i.e., the hypothesis with the highest likelihood) is simpler if the path toward this point is not obstructed by large "valleys" of very low likelihood. This is achieved by increasing low-likelihood hypotheses and decreasing high-likelihood hypotheses, thereby "flattening" the "knowledge landscape." This allows the learner to consistently move in the direction of the best hypothesis instead of "jumping" to the highest point after rejecting many low-likelihood hypotheses. A computational model by Ullman et al. (2012) showed how using these tempering ideas can account for how children and adults learn taxonomies and simple theories of magnetism.

A different approach to improve inference, which can be used alongside tempering, is motivated by the empirical insight that children and adults do not treat evidence they observe from other people's intentional actions as if it is sampled at random. Instead, children and adults exhibit an inductive bias to assume that others engage in behaviors that most efficiently achieve their goals and maximize their utilities (Gweon, 2021; Gweon et al., 2010; Jara-Ettinger, Gweon, Schulz, & Tenenbaum, 2016; Jara-Ettinger, Schulz, & Tenenbaum, 2020; Shafto, Goodman, & Frank, 2012; Shafto et al., 2014). For example, a person trying to teach a learner how to do something has likely chosen examples to learn from that will maximize the learner's ability to grasp the concept presented, meaning that a learner can consider only hypotheses for which the teacher's examples would be maximally informative as likely to be true (Shafto & Goodman, 2008). However, in cultures where direct instruction and teaching of children is less common, such as the Yucatec Maya, children learn to imitate novel tasks equally well whether they observe an adult or are directly instructed by an adult (Shneidman, Gaskins, & Woodward, 2016), and outperform U.S. children when observational learning is necessary (Correa-Chávez & Rogoff, 2009). This suggests that children not only infer others' goals and use this information to constrain their inferences about the evidence they receive but flexibly learn the appropriate strength of their inferences given the information environment they encounter in their daily lives.

The impact of these assumptions about how data are sampled was put to the test in a recent model of compositional visual concept learning tasks (Bongard problems; Depeweg, Rothkopf, & Jäkel, 2024). Bongard problems are challenging inference tasks, similar to Raven's progressive matrices, in which several visual examples of a concept are provided, and the learner has to infer the missing visual depiction. Depeweg et al. (2024) suggested a model that searched for solutions to these problems using formal language over visual concepts. However, to achieve a reasonable fit to human inferences, they had to introduce an additional assumption: that the examples provided in the Bongard tasks were not selected at random but were minimally informative. This assumption, similar to the sampling assumptions in Shafto and Goodman (2008), improved the models' ability to infer the correct concept. Depeweg and colleagues contrasted models with and without this pragmatic assumption and estimated that with the assumption, the correct solution was about 10% more likely than without it, considerably speeding up inference.

Another approach has sought to explicitly develop algorithmic-level theories (Marr, 1982) that capture the processes by which people revise their beliefs, including the role of domain-general cognitive resources such as working memory and other executive functions in inference. For example, one family of models proposes that the human mind only approximates probabilistic inference by sampling one or more hypotheses from mental simulation or memory (Chater et al., 2020; Sanborn & Chater, 2016; Sanborn, Yan, & Tsvetkov, 2024). While the concrete sampling approaches proposed as models of human inference differ, from simple procedures that only sample a new hypothesis when the previous one was unsuccessful (Bonawitz, Denison, Gopnik, & Griffiths, 2014) to more complex sampling paradigms from the applied statistics literature (e.g., Zhu, Sundh, Spicer, Chater, & Sanborn, 2024), the majority of these accounts assume a sequential, stochastic sampling process in which hypotheses are updated incrementally and locally (e.g., Bramley et al., 2017; Dasgupta, Schulz, & Gershman, 2017; Sanborn, Griffiths, & Navarro, 2010). This approach greatly simplifies inference, as only incremental changes to current hypotheses must be suggested, and only parts of the hypothesis are considered for revision in each step. However, these accounts also suggest signature limits of these processes, for instance, that the order in which evidence is received by learners may strongly impact what knowledge they acquire.

Some of the above sampling models (e.g., Sanborn et al., 2010) include a parameter for the number of simultaneous hypotheses being considered that can be as low as 1. When encountering conflicting evidence, a learner with one candidate hypothesis cannot simply lower the probability of this hypothesis but must revise the belief by generating an alternative hypothesis or, failing that, simply continue to believe their current one, which can result in characteristic "mode collapse" failures when two or more distinct possibilities must be considered simultaneously to succeed on a task.

This approach may explain why and how young children sometimes struggle to represent simultaneous alternative possibilities, for instance, when reasoning about future or counterfactual events or outcomes (e.g., Gerstenberg et al., 2021; Leahy & Carey, 2020). On these tasks, young children consistently put a hand under just one of a Y-maze's exit holes because they cannot represent coming out of the right hole and out of the left hole simultaneously. The capacity to simultaneously represent two or more hypotheses results in qualitatively different model predictions: Uncertainty can be represented not just as the "stickiness" of a candidate hypothesis but as a probability distribution over multiple candidate hypotheses. More generally, model parameters such as the number of simultaneous hypotheses being considered, the number of data points considered, and the depth of processing when revising one's beliefs could be applied to represent not only individual and situational differences in the cognitive resources that people apply to solving a task (e.g., Gelpí, León-Villagrá, Cunningham, Lucas, & Buchsbaum, 2025) but also to characterize the ongoing development of capacities like memory or executive function in children.

As described above, children and adults do not passively encounter information but actively accumulate evidence and construct their own learning curricula. In experiments with adults and children, Bramley and Xu (2023) showed that active search and explicit hypothesis generation patterns could be captured by a structured knowledge representation, where new hypotheses were derived by modifying previous inferences. Moreover, children had less systematic hypothesis generation and simplified less than adults, potentially due to more explorative goals. In a similar vein, models of adults' actions in a causal reasoning experiment showed that people mostly only consider hypotheses to account for the most recent evidence and plan actions using information about the anticipated reward but also taking into account the effort and time required to perform these actions (Bramley et al., 2017; Gong, Valentin, Lucas, & Bramley, 2024).

Finally, an agent's search for the appropriate hypotheses can be improved by not being exclusively driven by immediate utility. One candidate source of utility, beyond immediate reward, that has been suggested in cognitive science and artificial intelligence (AI) is intrinsic curiosity (Kidd & Hayden, 2015). These models posit that curiosity intrinsically increases the utility of certain actions in the present, anticipating that these actions will be important in the future (Dubey & Griffiths, 2020). Similar arguments have been put forward for action selection in causal settings, where an intrinsic utility of "empowerment," that is, a reward for bringing into effect one's planned actions, could guide an agent's search for causal hypotheses (Yiu, Allen, Ginosar, & Gopnik, 2026) as well as help provide structure in inherently unbounded, open-ended spaces (Chu & Schulz, 2022; Rule et al., 2020). Finally, assuming that other agents may also be driven by utilities that go beyond their immediate rewards may similarly improve our ability to learn both from and about them.

## 4. Conclusions, links to predictive processing, and questions for future work

From learning what kinds of materials are subject to invisible physical forces such as magnetism, or how visually similar animals such as dolphins and sharks are more distantly related than visually distinct animals such as mice and elephants, children face the daunting task of developing categorical and causal models that accurately capture a deeply complex world. Rational constructivism characterizes the process by which children and adults continuously revise and change their theories with a set of efficient inferential learning mechanisms in order to learn these models, despite the sparsity of the evidence available in the world.

Although developed principally as a model of cognitive development, rational constructivism is compatible with a broader set of theories within cognitive science inspired by Bayesian probabilistic modeling (e.g., Chater, Tenenbaum, & Yuille, 2006; Goodman et al., 2011; Ullman & Tenenbaum, 2020). Similarly, many of the commitments of rational constructivism also map onto elements of the predictive processing framework, the subject of the complementary introductory paper to this special issue (Clark, 2013; Friston, 2009; Friston et al., 2009; Sprevak & Smith, 2023). Beyond the centrality of Bayesian inference to both models—as a way to evaluate and revise hypotheses in rational constructivism and as a way to minimize internally generated prediction errors within predictive processing—both frameworks place a strong importance on the role of the learner in actively generating evidence. Rational constructivist learners use their own actions and interventions to obtain information and, in doing so, test their hypotheses about the world; a predictive processing agent engages in "active inference," predicting not just their own sensory inputs but their action plans based on the expected information gain and reward of the plan. Given that the theory of rational constructivism chiefly makes claims about children's and adults' learning and representation at the computational and algorithmic level of analysis, the similarities between rational constructivism and predictive processing suggest that combining insights from these two theories may offer a fruitful avenue of future research to answer the question of how the human brain can implement procedures for hypothesis generation, evaluation, and revision, as well as how learning in other domains, such as motor or procedural learning, develops across time.

One important question that computational models using structured representations face is what the innate building blocks used to assemble these representations consist of, and how they could be represented in the brain. This is especially challenging given that many of the models presented here assume sophisticated, domain-specific building blocks, such as physical forces or perceptual properties (color, mass, etc.). Piantadosi (2021) outlined how these domain-specific expressions could be learned without assuming cognitive primitives that are already meaningful. The formalism suggested by Piantadosi, combinatory logic, can learn these building blocks from more general primitives and the dynamics of the compositional rules. They showed that these primitives do not have meaning in themselves but can be learned to correspond to meaningful rules such as conjunction, disjunction, or implicature, which can be learned from domain-general internal representations. Future research should elaborate on this analysis, producing a more extensive account of how these formalisms can produce the structural representations that computational models commonly assume, as well as provide a closer examination of how the dynamics of representational learning implied by these approaches compare to the diverse forms that children's developing concepts take across cultures (e.g., Jara-Ettinger et al., 2022).

A related issue is the question of how these approaches can derive truly novel expressions. One concern is that theories are constructed from the combination of simpler, potentially innate building blocks, which will be too limited to account for the large conceptual leaps that children and adults can produce. This criticism seems particularly challenging given the emphasis of most computational models presented here, which assume that theory revision relies on small successive modifications of the learners' knowledge representation. One possible response to these criticisms is computational models that assume that learners can compress previously useful knowledge representations and store these in memory. These chunks can be composed of multiple simpler terms and allow the learner to build rich representations. As a result, learners can re-use these complex concepts to develop effective strategies for information seeking (Liquin et al., 2024) and potentially produce more radical theory revisions (Zhao et al., 2024). However, more research is required to assess if these dynamics of chunking and subsequent theory revision correspond to conceptual changes found in developmental studies. Specifically, we currently lack microgenetic data of children's conceptual change that provides an empirical testbed for these computational models.

The computational models presented in this paper provide precise and testable implementations of the theoretical commitments of rational constructivism. Therefore, they make predictions about how and under which circumstances concepts or theories can change in development. For example, Kemp and Tenenbaum (2008) showcased how a computational model of children's learning of animal categories could account for a shift from initially flat partitions of animals into disjoint groups to a tree-like taxonomic representation. The model produced this shift in purely data-driven form, predicting that given a critical number of encountered animal features a taxonomic organization should emerge.

Similarly, a model developed by Bonawitz et al. (2019) suggested a model learning a theory of magnetism that, given sufficient evidence, revised an early erroneous theory to produce a causally correct theory of magnetic forces. While these and other models predict how individual learners revise initially erroneous theories that account well for cross-sectional developmental experiments, few comparisons to individual learning trajectories have been conducted (but see Colantonio et al., 2022). Together these advances suggest that future research should focus on microgenetic or longitudinal experimental designs (Siegler, 2006) or "mini-microgenetic" studies (e.g., Bonawitz et al., 2014; Colantonio et al., 2022) that assess learning on every trial to test the theoretical predictions of these models.

Finally, all accounts presented here operate at the computational and algorithmic level of explanation (Marr, 1982). To present a full account of children's development and learning, we require a better theoretical understanding of how structured knowledge can be represented in the brain, how the brain can produce the incremental and local hypothesis revision suggested by these models, and how probabilistic inference can be performed with neural circuitry corresponding to these models. This challenge points to a long debate in the cognitive sciences, discussed at least since McClelland and Rumelhart (1985): Does cognition require discrete representations, such as symbols, rules, or concepts? Or can approaches that do not require symbols, and instead operate over distributed activity patterns, achieve similar expressive and inferential power? And while the work discussed in this paper showcases significant steps forward in addressing many longstanding debates in cognitive science, that we arrive back at this question suggests that there is much additional work to be done.

## Note

1 While outside the scope of this review, some even propose that natural language is the LOT (e.g., Asoulin, 2016; Hinzen, 2013).

## References


Asaba, M., Li, X., Yow, W. Q., & Gweon, H. (2022). Children selectively demonstrate their competence to a puppet when others depict it as an agent. *Cognitive Development*, *62*, 101186. https://doi.org/10.1016/j.cogdev.2022.101186

Asoulin, E. (2016). Language as an instrument of thought. *Glossa: A Journal of General Linguistics*, *1*(1), 1–23. https://doi.org/10.5334/gjgl.34

Austerweil, J. L., Gershman, S. J., & Griffiths, T. L. (2015). Structure and flexibility in Bayesian models of cognition. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds), *The Oxford handbook of computational and mathematical psychology* (pp. 187–208). Oxford, England: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199957996.013.9

Baillargeon, R. (2002). The acquisition of physical knowledge in infancy: A summary in eight lessons. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 47–83). Hoboken, NJ: John Wiley & Sons. https://doi.org/10.1002/9780470996652.ch3

Bardi, L., Regolin, L., & Simion, F. (2011). Biological motion preference in humans at birth: Role of dynamic and configural properties. *Developmental Science*, *14*(2), 353–359. https://doi.org/10.1111/j.1467-7687.2010.00985.x.

Bertenthal, B. I., Proffitt, D. R., & Cutting, J. E. (1984). Infant sensitivity to figural coherence in biomechanical motions. *Journal of Experimental Child Psychology*, *37*(2), 213–230. https://doi.org/10.1016/0022-0965(84)90001-8.

Bhui, R., Lai, L., & Gershman, S. J. (2021). Resource-rational decision making. *Current Opinion in Behavioral Sciences*, *41*, 15–21. https://doi.org/10.1016/j.cobeha.2021.02.015

Blake, R. (1993). Cats perceive biological motion. *Psychological Science*, *4*(1), 54–57. https://doi.org/10.1111/j.1467-9280.1993.tb00557.x

Blanco, N. J., & Sloutsky, V. M. (2021). Systematic exploration and uncertainty dominate young children's choices. *Developmental Science*, *24*(2), e13026. https://doi.org/10.1111/desc.13026.

Bonawitz, E. B., Denison, S., Gopnik, A., & Griffiths, T. L. (2014). Win-Stay, Lose-Sample: A simple sequential algorithm for approximating Bayesian inference. *Cognitive Psychology*, *74*, 35–65. https://doi.org/10.1016/j.cogpsych.2014.06.003

Bonawitz, E. B., & Lombrozo, T. (2012). Occam's rattle: Children's use of simplicity and probability to constrain inference. *Developmental Psychology*, *48*(4), 1156. https://doi.org/10.1037/a0026471.

Bonawitz, E. B., van Schijndel, T. J., Friel, D., & Schulz, L. (2012). Children balance theories and evidence in exploration, explanation, and learning. *Cognitive Psychology*, *64*(4), 215–234. https://doi.org/10.1016/j.cogpsych.2011.12.002.

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*(3), 322–330. https://doi.org/10.1016/j.cognition.2010.10.001.

Bonawitz, E., Ullman, T. D., Bridgers, S., Gopnik, A., & Tenenbaum, J. B. (2019). Sticking to the evidence? A behavioral and computational case study of micro-theory change in the domain of magnetism. *Cognitive Science*, *43*(8), e12765. https://doi.org/10.1111/cogs.12765.

Boole, G. (1854). *An investigation of the laws of thought: On which are founded the mathematical theories of logic and probabilities* (Vol. 2). London: Walton and Maberly.

Brainerd, C. J. (1978). The stage question in cognitive-developmental theory. *Behavioral and Brain Sciences*, *1*(2), 173–182. https://doi.org/10.1017/s0140525×00073842.

Bramley, N. R., Dayan, P., Griffiths, T. L., & Lagnado, D. A. (2017). Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*, *124*(3), 301–338. https://doi.org/10.1037/rev0000061.

Bramley, N. R., Jones, A., Gureckis, T. M., & Ruggeri, A. (2022). Children's failure to control variables may reflect adaptive decision-making. *Psychonomic Bulletin & Review*, *29*(6), 2314–2324. https://doi.org/10.3758/s13423-022-02120-1.

Bramley, N. R., & Xu, F. (2023). Active inductive inference in children and adults: A constructivist perspective. *Cognition*, *238*, 105471. https://doi.org/10.1016/j.cognition.2023.105471.

Brandone, A. C., Cimpian, A., Leslie, S. J., & Gelman, S. A. (2012). Do lions have manes? For children, generics are about kinds rather than quantities. *Child Development*, *83*(2), 423–433. https://doi.org/10.1111/j.1467-8624.2011.01708.x.

Bridgers, S., Buchsbaum, D., Seiver, E., Griffiths, T. L., & Gopnik, A. (2016). Children's causal inferences from conflicting testimony and observations. *Developmental Psychology*, *52*(1), 9–18. https://doi.org/10.1037/a0039830.

Brockbank, E., & Walker, C. M. (2022). Explanation impacts hypothesis generation, but not evaluation, during learning. *Cognition*, *225*, 105100. https://doi.org/10.1016/j.cognition.2022.105100.

Buchsbaum, D., Bridgers, S., Whalen, A., Seiver, E., Griffiths, T., & Gopnik, A. (2012). Do I know that you know what you know? Modeling testimony in causal inference. In *Proceedings of the annual meeting of the Cognitive Science Society*, Sapporo, Japan.

Buchsbaum, D., Gopnik, A., Griffiths, T. L., & Shafto, P. (2011). Children's imitation of causal action sequences is influenced by statistical and pedagogical evidence. *Cognition*, *120*(3), 331–340. https://doi.org/10.1016/j.cognition.2010.12.001.

Buchsbaum, D., Seiver, E., Bridgers, S., & Gopnik, A. (2012). Learning about causes from people and about people as causes: Probabilistic models and social causal reasoning. *Advances in Child Development and Behavior*, *43*, 125–160. https://doi.org/10.1016/B978-0-12-397919-3.00005-8

Carey, S. (1985). *Conceptual change in childhood*. Cambridge: MIT Press.

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

Carey, S., & Spelke, E. (1994). Domain-specific knowledge and conceptual change. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain Specificity in Cognition and Culture* (pp. 169–200). Cambridge University Press. https://doi.org/10.1017/CBO9780511752902.008

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, *10*(7), 287–291. https://doi.org/10.1016/j.tics.2006.05.007.

Chater, N., Zhu, J. Q., Spicer, J., Sundh, J., León-Villagrá, P., & Sanborn, A. (2020). Probabilistic biases meet the Bayesian brain. *Current Directions in Psychological Science*, *29*(5), 506–512. https://doi.org/10.1177/0963721420954801

Chu, J., & Schulz, L. E. (2020). Play, curiosity, and cognition. *Annual Review of Developmental Psychology*, *2*(1), 317–343. https://doi.org/10.1146/annurev-devpsych-070120-014806

Chu, J., & Schulz, L. (2022). "Because I want to": Valuing goals for their own sake. *Proceedings of the Annual Meeting of the Cognitive Science Society*, Toronto, Canada (pp. 1263–1269).

Cimpian, A., & Markman, E. M. (2009). Information learned from generic language becomes central to children's biological concepts: Evidence from their open-ended explanations. *Cognition*, *113*(1), 14–25. https://doi.org/10.1016/j.cognition.2009.07.004.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525×12000477.

Cohen, L. B., & Amsel, G. (1998). Precursors to infants' perception of the causality of a simple event. *Infant Behavior and Development*, *21*(4), 713–731. https://doi.org/10.1016/S0163-6383(98)90040-6

Colantonio, J. A., Bascandziev, I., Theobald, M., Brod, G., & Bonawitz, E. (2022). Priors, progressions, and predictions in science learning: theory-based Bayesian models of children's revising beliefs of water displacement. *IEEE Transactions on Cognitive and Developmental Systems*, *15*(3), 1487–1500. https://doi.org/10.1109/TCDS.2022.3220963

Correa-Chávez, M., & Rogoff, B. (2009). Children's attention to interactions directed to others: Guatemalan Mayan and European American patterns. *Developmental Psychology*, *45*(3), 630–641. https://doi.org/10.1037/a0014144.

Csibra, G., Bíró, S., Koós, O., & Gergely, G. (2003). One-year-old infants use teleological representations of actions productively. *Cognitive Science*, *27*(1), 111–133. https://doi.org/10.1207/s15516709cog2701_4

Dasgupta, I., Schulz, E., & Gershman, S. J. (2017). Where do hypotheses come from? *Cognitive Psychology*, *96*, 1–25. https://doi.org/10.1016/j.cogpsych.2017.05.001

Denison, S., Reed, C., & Xu, F. (2013). The emergence of probabilistic reasoning in very young infants: Evidence from 4.5- and 6-month-olds. *Developmental Psychology*, *49*(2), 243–249. https://doi.org/10.1037/a0028278.

Denison, S., & Xu, F. (2010). Twelve-to 14-month-old infants can predict single-event probability with large set sizes. *Developmental Science*, *13*(5), 798–803. https://doi.org/10.1111/j.1467-7687.2009.00943.x.

Depeweg, S., Rothkopf, C. A., & Jäkel, F. (2024). Solving Bongard problems with a visual language and pragmatic constraints. *Cognitive Science*, *48*(5), e13432. https://doi.org/10.1111/cogs.13432.

Dewar, K. M., & Xu, F. (2010). Induction, overhypothesis, and the origin of abstract knowledge: Evidence from 9-month-old infants. *Psychological Science*, *21*(12), 1871–1877. https://doi.org/10.1177/0956797610388810.

Diesendruck, G. (2003). Categories for names or names for categories?The interplay between domain-specific conceptual structure and language. *Language and Cognitive Processes*, *18*(5–6), 759–787. https://doi.org/10.1080/01690960344000116

Doan, T., Denison, S., Lucas, C., & Gopnik, A. (2015). Learning to reason about desires: An infant training study. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th annual meeting of the Cognitive Science Society* (pp. 578–583). Austin, TX: Cognitive Science Society.

Doan, T., Friedman, O., & Denison, S. (2021). Toddlers and preschoolers understand that some preferences are more subjective than others. *Child Development*, *92*(3), 853–861. https://doi.org/10.1111/cdev.13581.

Dragon, M., & Poulin-Dubois, D. (2023). To copy or not to copy: A comparison of selective trust and overimitation in young children. *Cognitive Development*, *66*, 101316. https://doi.org/10.1016/j.cogdev.2023.101316

Dubey, R., & Griffiths, T. L. (2020). Understanding exploration in humans and machines by formalizing the function of curiosity. *Current Opinion in Behavioral Sciences*, *35*, 118–124. https://doi.org/10.1016/j.cobeha.2020.07.008

Ellis, K., Wong, C., Nye, M., Sablé-Meyer, M., Morales, L., Hewitt, L., Cary, L., Solar-Lezama, A., & Tenenbaum, J. B. (2021). DreamCoder: Bootstrapping inductive program synthesis with wake-sleep library learning. *Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation*, 835–850. https://doi.org/10.1145/3453483.3454080.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/5929.001.0001

Fedyk, M., & Xu, F. (2018). The epistemology of rational constructivism. *Review of Philosophy and Psychology*, *9*, 343–362. https://doi.org/10.1007/s13164-017-0372-1

Felsche, E., Stevens, P., Völter, C. J., Buchsbaum, D., & Seed, A. M. (2023). Evidence for abstract representations in children but not capuchin monkeys. *Cognitive Psychology*, *140*, 101530. https://doi.org/10.1016/j.cogpsych.2022.101530.

Felsche, E., Völter, C. J., Herrmann, E., Seed, A. M., & Buchsbaum, D. (2024). How can I find what I want? Can children, chimpanzees and capuchin monkeys form abstract representations to guide their behavior in a sampling task? *Cognition*, *245*, 105721. https://doi.org/10.1016/j.cognition.2024.105721.

Fodor, J. A. (1975). *The language of thought* (Vol. 5). Cambridge, MA: Harvard University Press.

Friston, K. (2009). The free-energy principle: A rough guide to the brain?. *Trends in Cognitive Sciences*, *13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005.

Friston, K., & Kiebel, S. J. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society of London, Series B*, *364*(1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300

Ganea, P. A., Larsen, N. E., & Venkadasalam, V. P. (2021). The role of alternative theories and anomalous evidence in children's scientific belief revision. *Child Development*, *92*(3), 1137–1153. https://doi.org/10.1111/cdev.13481.

Garnelo, M., & Shanahan, M. (2019). Reconciling deep learning with symbolic artificial intelligence: Representing objects and relations. *Current Opinion in Behavioral Sciences*, *29*, 17–23. https://doi.org/10.1016/j.cobeha.2018.12.010

Gelman, S. A., & Davidson, N. S. (2013). Conceptual influences on category-based induction. *Cognitive Psychology*, *66*(3), 327–353. https://doi.org/10.1016/j.cogpsych.2013.02.001.

Gelman, S. A., Goetz, P. J., Sarnecka, B. W., & Flukes, J. (2008). Generic language in parent-child conversations. *Language Learning and Development*, *4*(1), 1–31. https://doi.org/10.1080/15475440701542625.

Gelman, S. A., Leslie, S. J., Was, A. M., & Koch, C. M. (2015). Children's interpretations of general quantifiers, specific quantifiers and generics. *Language, Cognition and Neuroscience*, *30*(4), 448–461. https://doi.org/10.1080/23273798.2014.931591.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, *23*(3), 183–209. https://doi.org/10.1016/0010-0277(86)90034-X.

Gelman, S. A., & Raman, L. (2003). Preschool children use linguistic form class and pragmatic cues to interpret generics. *Child Development*, *74*(1), 308–325. https://doi.org/10.1111/1467-8624.00537.

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.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the non-obvious. *Cognition*, *38*(3), 213–244. https://doi.org/10.1016/0010-0277(91)90007-Q.

Gelman, S. A., Wilcox, S. A., & Clark, E. V. (1989). Conceptual and lexical hierarchies in young children. *Cognitive Development*, *4*(4), 309–326. https://doi.org/10.1016/S0885-2014(89)90022-1

Gelpí, R. A., & Buchsbaum, D. (2025). Children as cultural explorers: How Imitation, pedagogy, and selective trust prepare children for learning in the cultural niche. In J. J. Tehrani, J. Kendal, & R. L. Kendal (Eds.), *The Oxford handbook of cultural evolution* (pp. 241–253). Oxford, England: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198869252.013.19

Gelpí, R. A., León-Villagrá, P., Cunningham, W. A., Lucas, C. G., & Buchsbaum, D. (2025). Resource-rational belief revision can mitigate as well as amplify polarization. In A. Ruggeri, D. Barner, C. Walker, & N. Bramley (Eds.), *Proceedings of the annual meeting of the Cognitive Science Society* (Vol. 47, pp. 767–774). Austin, TX: Cognitive Science Society.

Gelpí, R. A., Otsubo, K., Whalen, A., & Buchsbaum, D. (2025). Investigating sensitivity to shared information and personal experience in children's use of majority information. *Open Mind*, *9*, 240–265. https://doi.org/10.1162/opmi_a_00182

Gelpí, R. A., Whalen, A., Griffiths, T. L., Xu, F., & Buchsbaum, D. (2025). Can Children and adults balance majority size with information quality in learning from preferences? *Journal of Experimental Psychology: General*, *154*(5), 1388–1406. https://doi.org/10.1037/xge0001724

Gelpi, R., Prystawski, B., Lucas, C. G., & Buchsbaum, D. (2020). Incremental Hypothesis Revision in Causal Reasoning Across Development. Proceedings of the Annual Meeting of the Cognitive Science Society, *42*(0). https://escholarship.org/uc/item/78j348s3

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature*, *415*(6873), 755–755. https://doi.org/10.1038/415755a.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287–292. https://doi.org/10.1016/s0010-0277(97)00004-8.

Gerken, L. (2010). Infants use rational decision criteria for choosing among models of their input. *Cognition*, *115*(2), 362–366. https://doi.org/10.1016/j.cognition.2010.01.006.

Gershman, S. J. (2017). On the blessing of abstraction. *Quarterly Journal of Experimental Psychology*, *70*(3), 361–365. https://doi.org/10.1080/17470218.2016.1159706

Gerstenberg, T., & Stephan, S. (2021). A counterfactual simulation model of causation by omission. *Cognition*, *216*, 104842, https://doi.org/10.1016/j.cognition.2021.104842.

Giron, A. P., Ciranka, S., Schulz, E., van den Bos, W., Ruggeri, A., Meder, B., & Wu, C. M. (2023). Developmental changes in exploration resemble stochastic optimization. *Nature Human Behaviour*, *7*(11), 1955–1967. https://doi.org/10.1038/s41562-023-01662-1.

Goddu, M. K., & Gopnik, A. (2024). The development of human causal learning and reasoning. *Nature Reviews Psychology*, *3*, 319–339. https://doi.org/10.1038/s44159-024-00300-5

Gong, T., Valentin, S., Lucas, C., & Bramley, N. R. (2024). Paradoxical parsimony: How latent complexity favors theory simplicity. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Rotterdam, The Netherlands.

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, *32*(1), 108–154. https://doi.org/10.1080/03640210701802071.

Goodman, N. D., Ullman, T. D., & Tenenbaum, J. B. (2011). Learning a theory of causality. *Psychological Review*, *118*(1), 110–119. https://doi.org/10.1037/a0021336.

Goodman, N. (1955). *Fact, fiction, & forecast*. Cambridge: Harvard University Press.

Gopnik, A. (2020). Childhood as a solution to explore–exploit tensions. *Philosophical Transactions of the Royal Society*, *375*(1803), 20190502. https://doi.org/10.1098/rstb.2019.0502

Gopnik, A., & Schulz, L. (2004). Mechanisms of theory formation in young children. *Trends in Cognitive Sciences*, *8*(8), 371–377. https://doi.org/10.1016/j.tics.2004.06.005.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, *138*(6), 1085–1108. https://doi.org/10.1037/a0028044.

Graham, S. A., Gelman, S. A., & Clarke, J. (2016). Generics license 30-month-olds' inferences about the atypical properties of novel kinds. *Developmental Psychology*, *52*(9), 1353. https://doi.org/10.1037/dev0000183.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357–364. https://doi.org/10.1016/j.tics.2010.05.004.

Griffiths, T. L., Sobel, D. M., Tenenbaum, J. B., & Gopnik, A. (2011). Bayes and blickets: Effects of knowledge on causal induction in children and adults. *Cognitive Science*, *35*(8), 1407–1455. https://doi.org/10.1111/j.1551-6709.2011.01203.x.

Grigoroglou, M., & Ganea, P. A. (2022). Language as a mechanism for reasoning about possibilities. *Philosophical Transactions of the Royal Society B*, *377*(1866), 20210334. https://doi.org/10.1098/rstb.2021.0334

Gweon, H. (2021). Inferential social learning: Cognitive foundations of human social learning and teaching. *Trends in Cognitive Sciences*, *25*(10), 896–910. https://doi.org/10.1016/j.tics.2021.07.008.

Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences*, *107*(20), 9066–9071. https://doi.org/10.1073/pnas.1003095107

Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology*, *69*(1), 251–273. https://doi.org/10.1146/annurev-psych-122216-011710.

Hinzen, W. (2013). Narrow syntax and the language of thought. *Philosophical Psychology*, *26*(1), 1–23. https://doi.org/10.1080/09515089.2011.627537

Imai, M., & Gentner, D. (1997). A cross-linguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition*, *62*(2), 169–200. https://doi.org/10.1016/s0010-0277(96)00784-6.

Inagaki, K., & Hatano, G. (2004). Vitalistic causality in young children's naive biology. *Trends in Cognitive Sciences*, *8*(8), 356–362. https://doi.org/10.1016/j.tics.2004.06.004.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, *20*(8), 589–604. https://doi.org/10.1016/j.tics.2016.05.011.

Jara-Ettinger, J., Levy, R., Sakel, J., Huanca, T., & Gibson, E. (2022). The origins of the shape bias: Evidence from the Tsimane. *Journal of Experimental Psychology: General*, *151*(10), 2437–2447. https://doi.org/10.1037/xge0001195.

Jara-Ettinger, J., Schulz, L. E., & Tenenbaum, J. B. (2020). The naive utility calculus as a unified, quantitative framework for action understanding. *Cognitive Psychology*, *123*, 101334. https://doi.org/10.1016/j.cogpsych.2020.101334.

Jones, A., Swaboda, N., & Ruggeri, A. (2020). Developmental changes in question-asking. In K. H. Corriveau, L. P. Butler, & S. Ronfard (Eds.), *The questioning child: Insights from psychology & education* (pp. 118–143). Cambridge University Press. https://doi.org/10.1017/9781108553803.007

Keil, F. C., Levin, D. T., Richman, B. A., & Gutheil, G. (1999). Mechanism and explanation in the development of biological thought: The case of disease. In D. L. Medin & S. Atran (Eds.), *Folkbiology* (pp. 285–319). Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/3042.003.0010

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*(3), 307–321. https://doi.org/10.1111/j.1467-7687.2007.00585.x.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences*, *105*(31), 10687–10692. https://doi.org/10.1073/pnas.0802631105

Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review*, *116*(1), 20–58. https://doi.org/10.1037/a0014282.

Khemlani, S., Leslie, S. J., & Glucksberg, S. (2012). Inferences about members of kinds: The generics hypothesis. *Language and Cognitive Processes*, *27*(6), 887–900. https://doi.org/10.1080/01690965.2011.601900

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron*, *88*(3), 449–460. https://doi.org/10.1016/j.neuron.2015.09.010.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE*, *7*(5), e36399. https://doi.org/10.1371/journal.pone.0036399.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child Development*, *85*(5), 1795–1804. https://doi.org/10.1111/cdev.12263.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42. https://doi.org/10.1016/S0010-0277(02)00004-5.

Kushnir, T., & Gopnik, A. (2005). Young children infer causal strength from probabilities and interventions. *Psychological Science*, *16*(9), 678–683. https://doi.org/10.1111/j.1467-9280.2005.01595.x.

Kushnir, T., Xu, F., & Wellman, H. M. (2010). Young children use statistical sampling to infer the preferences of other people. *Psychological Science*, *21*(8), 1134–1140. https://doi.org/10.1177/0956797610376652.

Lake, B. M., Lawrence, N. D., & Tenenbaum, J. B. (2018). The emergence of organizing structure in conceptual representation. *Cognitive Science*, *42*, 809–832. https://doi.org/10.1111/cogs.12580.

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338. https://doi.org/10.1126/science.aab3050

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*(3), 299–321. https://doi.org/10.1016/0885-2014(88)90014-7

Lapidow, E., & Walker, C. M. (2020). Informative experimentation in intuitive science: Children select and learn from their own causal interventions. *Cognition*, *201*, 104315. https://doi.org/10.1016/j.cognition.2020.104315.

Laurence, S., & Margolis, E. (2024). *The building blocks of thought: A rationalist account of the origins of concepts*. Oxford, England: Oxford University Press. https://doi.org/10.1093/9780191925375.001.0001

Leahy, B. P., & Carey, S. E. (2020). The Acquisition of Modal Concepts. *Trends in Cognitived Sciences*, *24*(1), 65–78. https://doi.org/10.1016/j.tics.2019.11.004

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

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, e1. https://doi.org/10.1017/s0140525×1900061x

Liquin, E. G., Rhodes, M., & Gureckis, T. M. (2024). Reuse and remixing in question asking across development. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Rotterdam, The Netherlands.

Lombrozo, T. (2020). Learning by thinking in science and in everyday life. In P. Godfrey-Smith & A. Levy (Eds.), *The scientific imagination* (pp. 230–249). New York, NY: Oxford University Press. https://doi.org/10.1093/oso/9780190212308.003.0010

Lucas, C. G., Griffiths, T. L., Xu, F., Fawcett, C., Gopnik, A., Kushnir, T., Markson, L., & Hu, J. (2014). The child as econometrician: A rational model of preference understanding in children. *PLoS ONE*, *9*(3), e92160. https://doi.org/10.1371/journal.pone.0092160

Lucy, J. A., & Gaskins, S. (2001). Grammatical categories and the development of classification preferences: A comparative approach. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual devel-*

*opment* (pp. 257–283). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511620669.011

Ma, L., & Xu, F. (2011). Young children's use of statistical sampling evidence to infer the subjectivity of preferences. *Cognition*, *120*(3), 403–411. https://doi.org/10.1016/j.cognition.2011.02.003.

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge: MIT Press. https://doi.org/10.7551/mitpress/1750.001.0001

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: WH Freeman. https://doi.org/10.7551/mitpress/9780262514620.001.0001

McClelland, J. L., & Rumelhart, D. E. (1985). Level's indeed! A response to Broadbent. *Journal of Experimental Psychology: General*, *114*, 193–197. https://psycnet.apa.org/doi/10.1037/0096-3445.114.2.193

Meder, B., Wu, C. M., Schulz, E., & Ruggeri, A. (2021). Development of directed and random exploration in children. *Developmental Science*, *24*(4), e13095. https://doi.org/10.1111/desc.13095.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*(5), 838–850. https://doi.org/10.1037/0012-1649.31.5.838.

Mills, C. M. (2013). Knowing when to doubt: Developing a critical stance when learning from others. *Developmental Psychology*, *49*(3), 404–418. https://doi.org/10.1037/a0029500.

Muentener, P., & Bonawitz, E. B. (2017). The development of causal reasoning. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 677–698). New York, NY: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199399550.013.40

Munakata, Y., & McClelland, J. L. (2003). Connectionist models of development. *Developmental Science*, *6*(4), 413–429. https://doi.org/10.1111/1467-7687.00296

Nussenbaum, K., & Hartley, C. A. (2019). Reinforcement learning across development: What insights can we draw from a decade of research? *Developmental Cognitive Neuroscience*, *40*, 100733. https://doi.org/10.1016/j.dcn.2019.100733.

Oakes, L. M., Madole, K. L., & Cohen, L. B. (1991). Infants' object examining: Habituation and categorization. *Cognitive Development*, *6*(4), 377–392. https://doi.org/10.1016/0885-2014(91)90045-F

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oSo/9780198524496.001.0001

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, *82*(4), 669–688. https://doi.org/10.1093/biomet/82.4.669

Pelz, M., & Kidd, C. (2020). The elaboration of exploratory play. *Philosophical Transactions of the Royal Society*, *375*(1803), 20190503. https://doi.org/10.1098/rstb.2019.0503

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, *120*(3), 302–321. https://doi.org/10.1016/j.cognition.2010.11.015

Pham, T., & Buchsbaum, D. (2020). Children's use of majority information is influenced by pragmatic inferences and task domain. *Developmental Psychology*, *56*(2), 312–323. https://doi.org/10.1037/dev0000857

Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books. https://psycnet.apa.org/doi/10.1037/11168-000

Piantadosi, S. T. (2021). The computational origin of representation. *Minds and Machines*, *31*(1), 1–58. https://doi.org/10.1007/s11023-020-09540-9.

Piantadosi, S. T., Jara-Ettinger, J., & Gibson, E. (2014). Children's learning of number words in an indigenous farming-foraging group. *Developmental Science*, *17*(4), 553–563. https://doi.org/10.1111/desc.12078.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, *306*(5695), 499–503. https://doi.org/10.1126/science.1102085.

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology*, *33*(1), 12–21. https://doi.org/10.1037/0012-1649.33.1.12.

Ritter, S., Barrett, D. G., Santoro, A., & Botvinick, M. M. (2017). Cognitive psychology for deep neural networks: A shape bias case study. *International Conference on Machine Learning*, Sydney, Australia (pp. 2940–2949).

Ruggeri, A., & Lombrozo, T. (2015). Children adapt their questions to achieve efficient search. *Cognition*, *143*, 203–216. https://doi.org/10.1016/j.cognition.2015.07.004.

Ruggeri, A., Sim, Z. L., & Xu, F. (2017). Why is Toma late to school again? Preschoolers identify the most informative questions. *Developmental Psychology*, *53*(9), 1620–1632. https://doi.org/10.1037/dev0000340.

Ruggeri, A., Stanciu, O., Pelz, M., Gopnik, A., & Schulz, E. (2024). Preschoolers search longer when there is more information to be gained. *Developmental Science*, *27*(1), e13411. https://doi.org/10.1111/desc.13411.

Ruggeri, A., Swaboda, N., Sim, Z. L., & Gopnik, A. (2019). Shake it baby, but only when needed: Preschoolers adapt their exploratory strategies to the information structure of the task. *Cognition*, *193*, 104013. https://doi.org/10.1016/j.cognition.2019.104013.

Rule, J. S., Tenenbaum, J. B., & Piantadosi, S. T. (2020). The child as hacker. *Trends in Cognitive Sciences*, *24*(11), 900–915. https://doi.org/10.1016/j.tics.2020.07.005.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926.

Samet, J., Bronstein, D., Brown, D., Helmig, C., & Normore, C. (2024). The historical controversies surrounding innateness. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy*. Stanford: Stanford University. https://plato.stanford.edu/archives/sum2024/entries/innateness-history/

Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences*, *20*(12), 883–893. https://doi.org/10.1016/j.tics.2016.10.003.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, *117*(4), 1144–1167. https://doi.org/10.1037/a0020511.

Sanborn, A. N., Yan, H., & Tsvetkov, C. (2024). Combining meta-learned models with process models of cognition. *Behavioral and Brain Sciences*, *47*, e163. https://doi.org/10.1017/s0140525x24000165

Saxe, R., & Carey, S. (2006). The perception of causality in infancy. *Acta Psychologica*, *123*(1-2), 144–165. https://doi.org/10.1016/j.actpsy.2006.05.005

Schluessel, V., Kortekamp, N., Cortes, J. O., Klein, A., & Bleckmann, H. (2015). Perception and discrimination of movement and biological motion patterns in fish. *Animal Cognition*, *18*, 1077–1091. https://doi.org/10.1007/s10071-015-0876-y.

Schulz, E., Wu, C. M., Ruggeri, A., & Meder, B. (2019). Searching for rewards like a child means less generalization and more directed exploration. *Psychological Science*, *30*(11), 1561–1572. https://doi.org/10.1177/0956797619863663.

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology*, *43*(4), 1045–1050. https://doi.org/10.1037/0012-1649.43.4.1045.

Schulz, L. E., Standing, H. R., & Bonawitz, E. B. (2008). Word, thought, and deed: The role of object categories in children's inductive inferences and exploratory play. *Developmental Psychology*, *44*(5), 1266–1276. https://doi.org/10.1037/0012-1649.44.5.1266.

Schulz, L. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in Cognitive Sciences*, *16*(7), 382–389. https://doi.org/10.1016/j.tics.2012.06.004.

Schwier, C., van Maanen, C., Carpenter, M., & Tomasello, M. (2006). Rational imitation in 12-month-old infants. *Infancy*, *10*, 303–311. https://doi.org/10.1207/s15327078in1003_6

Schlottmann, A., & Ray, E. (2010). Goal attribution to schematic animals: do 6-month-olds perceive biological motion as animate? *Developmental Science*, *13*(1), 1–10. https://doi.org/10.1111/j.1467-7687.2009.00854.x

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

Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology*, *49*(1), 31–57. https://doi.org/10.1016/0022-0965(90)90048-D

Sebastián-Enesco, C., Guerrero, S., & Enesco, I. (2020). What makes children defy their peers? Chinese and Spanish preschoolers' decisions to trust (or not) peer consensus. *Social Development*, *29*(2), 494–508. https://doi.org/10.1111/sode.12416

Shafto, P., Goodman, N. D., & Frank, M. C. (2012). Learning from others: The consequences of psychological reasoning for human learning. *Perspectives on Psychological Science*, *7*(4), 341–351. https://doi.org/10.1177/1745691612448481.

Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology*, *71*, 55–89. https://doi.org/10.1016/j.cogpsych.2013.12.004.

Shafto, P., & Goodman, N. (2008). Teaching games: Statistical sampling assumptions for learning in pedagogical situations. In B. Love, K. McRae, & V. Sloutsky (Eds.), *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 30, pp. 1632–1637). Austin, TX: Cognitive Science Society.

Shneidman, L., Gaskins, S., & Woodward, A. (2016). Child-directed teaching and social learning at 18 months of age: Evidence from Yucatec Mayan and US infants. *Developmental Science*, *19*(3), 372–381. https://doi.org/10.1111/desc.12318.

Siegler, R. S. (2006). Microgenetic analyses of learning. In D. Kuhn, R. S. Siegler, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Cognition, perception, and language* (6th ed., pp. 464–510). Hoboken, NJ: Wiley. https://doi.org/10.1002/9780470147658.chpsy0211

Sim, Z. L., & Xu, F. (2017). Infants preferentially approach and explore the unexpected. *British Journal of Developmental Psychology*, *35*(4), 596–608. https://doi.org/10.1111/bjdp.12198

Simion, F., Leo, I., Turati, C., Valenza, E., & Dalla Barba, B. (2007). How face specialization emerges in the first months of life. *Progress in Brain Research*, *164*, 169–185. https://doi.org/10.1016/S0079-6123(07)64009-6

Sobel, D. M., & Kirkham, N. Z. (2007). Bayes nets and babies: Infants' developing statistical reasoning abilities and their representation of causal knowledge. *Developmental Science*, *10*(3), 298–306. https://doi.org/10.1111/j.1467-7687.2007.00589.x

Sobel, D. M., & Sommerville, J. A. (2010). The importance of discovery in children's causal learning from interventions. *Frontiers in Psychology*, *1*, 176. https://doi.org/10.3389/fpsyg.2010.00176.

Spelke, E. S. (2022). *What babies know: Core Knowledge and Composition volume 1* (Vol. 1). Oxford, England: Oxford University Press. https://doi.org/10.1093/oso/9780190618247.001.0001

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96. https://doi.org/10.1111/j.1467-7687.2007.00569.x

Sprevak, M., & Smith, R. (2023). An introduction to predictive processing models of perception and decision-making. *Topics in Cognitive Science*, 1–28. https://doi.org/10.1111/tops.12704

Sumner, E., Steyvers, M., & Sarnecka, B. W. (2019). It's not the treasure, it's the hunt: Children are more explorative on an explore/exploit task than adults. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *41*, Montreal, Canada. https://escholarship.org/uc/item/61r509wq

Tecwyn, E. C., Mazumder, P., & Buchsbaum, D. (2023). One- and two-year-olds grasp that causes must precede their effects. *Developmental Psychology*, *59*(8), 1519–1531. https://doi.org/10.1037/dev0001551.

Teglas, E., Girotto, V., Gonzalez, M., & Bonatti, L. (2007). Intuitions of probabilities shape expectations about the future at 12 months and beyond. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 19156–19159. https://doi.org/10.1073/pnas.0700271104.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285. https://doi.org/10.1126/science.1192788.

Tessler, M. H., & Goodman, N. D. (2019). The language of generalization. *Psychological Review*, *126*(3), 395–436. https://doi.org/10.1037/rev0000142.

Ullman, T. D., Goodman, N. D., & Tenenbaum, J. B. (2012). Theory learning as stochastic search in the language of thought. *Cognitive Development*, *27*(4), 455–480. https://doi.org/10.1016/j.cogdev.2012.07.005

Ullman, T. D., Stuhlmüller, A., Goodman, N. D., & Tenenbaum, J. B. (2018). Learning physical parameters from dynamic scenes. *Cognitive Psychology*, *104*, 57–82. https://doi.org/10.1016/j.cogpsych.2017.05.006.

Ullman, T. D., & Tenenbaum, J. B. (2020). Bayesian models of conceptual development: Learning as building models of the world. *Annual Review of Developmental Psychology*, *2*(1), 533–558. https://doi.org/10.1146/annurev-devpsych-121318-084833

Vallortigara, G., Regolin, L., & Marconato, F. (2005). Visually inexperienced chicks exhibit spontaneous preference for biological motion patterns. *PLoS Biology*, *3*(7), e208. https://doi.org/10.1371/journal.pbio.0030208.

Walker, C. M., Bonawitz, E., & Lombrozo, T. (2017). Effects of explaining on children's preference for simpler hypotheses. *Psychonomic Bulletin & Review*, *24*, 1538–1547. https://doi.org/10.3758/s13423-016-1144-0.

Waxman, S. R., Lynch, E. B., Casey, K. L., & Baer, L. (1997). Setters and samoyeds: The emergence of subordinate level categories as a basis for inductive inference in preschool-age children. *Developmental Psychology*, *33*(6), 1074–1090. https://doi.org/10.1037/0012-1649.33.6.1074.

Waxman, S., & Gelman, R. (1986). Preschoolers' use of superordinate relations in classification and language. *Cognitive Development*, *1*(2), 139–156. https://doi.org/10.1016/S0885-2014(86)80016-8

Wellman, H. M., & Gelman, S. A. (1992). Cognitive development: foundational theories of core domains. *Annual Review of Psychology*. https://doi.org/10.1146/annurev.ps.43.020192.002005.

Woodward, A. L., & Sommerville, J. A. (2000). Twelve-month-old infants interpret action in context. *Psychological Science*, *11*(1), 73–77. https://doi.org/10.1111/1467-9280.00218.

Woodward, A. L., Sommerville, J. A., & Guajardo, J. J. (2001). How infants make sense of intentional action. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 149–171). Cambridge: MIT Press.

Xu, F. (2019). Towards a rational constructivist theory of cognitive development. *Psychological Review*, *126*(6), 841. https://doi.org/10.1037/rev0000153.

Xu, F. (2016). Preliminary thoughts on a rational constructivist approach to cognitive development: primitives, symbols, learning, and thinking. In D. Barner & A. Baron (Eds.), *Core knowledge and conceptual change* (pp. 11–28). New York: Oxford University Press. https://doi.org/10.1093/acprof:oSo/9780190467630.003.0002

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive Psychology*, *30*(2), 111–153. https://doi.org/10.1006/cogp.1996.0005.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences*, *105*(13), 5012–5015. https://doi.org/10.1073/pnas.0704450105

Xu, F., & Griffiths, T. L. (2011). Probabilistic models of cognitive development: Towards a rational constructivist approach to the study of learning and development. *Cognition*, *120*(3), 299–301. https://doi.org/10.1016/j.cognition.2011.06.008.

Xu, F., & Kushnir, T. (2013). Infants are rational constructivist learners. *Current Directions in Psychological Science*, *22*(1), 28–32. https://doi.org/10.1177/0963721412469396

Xu, F., & Kushnir, T. (2012). Preface to rational constructivism in cognitive development. In F. Xu & T. Kushnir (Eds.), *Advances in child development and behavior*, *43* (pp. xi–xiv). Cambridge: Academic Press. https://doi.org/10.1016/b978-0-12-397919-3.22001-7

Xu, F. (2007). Rational statistical inference and cognitive development. In P. Carruthers & S. Laurence (Eds.), *The innate mind: Foundations and the future* (pp. 199–215). Oxford, England: Oxford Academic. https://doi.org10.1093/acprof:oSo/9780195332834.003.0010

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245–272. https://doi.org/10.1037/0033-295X.114.2.245.

Yiu, E., Allen, K., Ginosar, S., & Gopnik, A. (2026). Empowerment gain and causal model construction: Children and adults are sensitive to controllability and variability in their causal interventions. *Philosophical Transactions of the Royal Society A*, *384*, 20250003. https://doi.org/10.1098/rsta.2025.0003.

Yu, C.-L., & Wellman, H. M. (2022). Young children treat puppets and dolls like real persons in theory of mind research: A meta-analysis of false-belief understanding across ages and countries. *Cognitive Development*, *63*, 101197. https://doi.org/10.1016/j.cogdev.2022.101197

Zhao, B., Bramley, N. R., & Lucas, C. (2022). Powering up causal generalization: A model of human conceptual bootstrapping with adaptor grammars. In J. Culbertson, A. Perfors, H. Rabagliati, & V. Ramenzoni (Eds.), *Proceedings of the 44th annual meeting of the Cognitive Science Society* (pp. 1819–1826). Austin, TX: Cognitive Science Society.

Zhao, B., Lucas, C. G., & Bramley, N. R. (2024). A model of conceptual bootstrapping in human cognition. *Nature Human Behaviour*, *8*(1), 125–136. https://doi.org/10.1038/s41562-023-01719-1.

Zhu, J. Q., Sundh, J., Spicer, J., Chater, N., & Sanborn, A. N. (2024). The autocorrelated Bayesian sampler: A rational process for probability judgments, estimates, confidence intervals, choices, confidence judgments, and response times. *Psychological Review*, *131*(2), 456. https://doi.org/10.1037/rev0000427.