Cognitive Psychology 104 (2018) 1–28

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

The role of domain-general cognitive resources in children's construction of a vitalist theory of biology

Igor Bascandziev[a,b,*], Nathan Tardiff[c], Deborah Zaitchik[a,d,e], Susan Carey[a]

[a] Department of Psychology, Harvard University, Cambridge, MA, United States
[b] Department of Psychology, Reed College, Portland, OR, United States
[c] Department of Psychology, University of Pennsylvania, Philadelphia, PA, United States
[d] Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States
[e] Department of Psychiatry, Harvard Medical School, Boston, MA, United States

ARTICLE INFO

Keywords:
Conceptual development
Executive function
Vitalism

ABSTRACT

Some episodes of learning are easier than others. Preschoolers can learn certain facts, such as "my grandmother gave me this purse," only after one or two exposures (easy to learn; fast mapping), but they require several years to learn that plants are alive or that the sun is not alive (hard to learn). One difference between the two kinds of knowledge acquisition is that hard cases often require conceptual construction, such as the construction of the biological concept *alive*, whereas easy cases merely involve forming new beliefs formulated over concepts the child already has (belief revision, a form of knowledge enrichment). We asked whether different domain-general cognitive resources support these two types of knowledge acquisition (conceptual construction and knowledge enrichment that supports fast mapping) by testing 82 6-year-olds in a pre-training/training/post-training study. We measured children's improvement in an episode involving theory construction (the beginning steps of acquisition of the framework theory of vitalist biology, which requires conceptual change) and in an episode involving knowledge enrichment alone (acquisition of little known facts about animals, such as the location of crickets' ears and the color of octopus blood). In addition, we measured children's executive functions and receptive vocabulary to directly compare the resources drawn upon in the two episodes of learning. We replicated and extended previous findings highlighting the differences between conceptual construction and knowledge enrichment, and we found that Executive Functions predict improvement on the Vitalism battery but not on the Fun Facts battery and that Receptive Vocabulary predicts improvement the Fun Facts battery but not on the Vitalism battery. This double dissociation provides new evidence for the distinction between the two types of knowledge acquisition, and bears on the nature of the learning mechanisms involved in each.

1. Introduction

Some episodes of knowledge acquisition are extraordinarily easy and some are extraordinarily hard. For example, adding new facts to our database on first encountering evidence for them, through testimony or observation, and retaining those facts for weeks or more is easy (Markson & Bloom, 1997). Conversely, adding new facts that express propositions central to a framework theory (Wellman & Gelman, 1992) that is not yet constructed by the child is very hard (Carey, 1985b, 2009). Many factors differentiate the

* Corresponding author at: Department of Psychology, Reed College, 3203 SE Woodstock Blvd, Portland, OR 97202, United States.
E-mail address: igb078@mail.harvard.edu (I. Bascandziev).

https://doi.org/10.1016/j.cogpsych.2018.03.002
Accepted 16 March 2018
Available online 26 March 2018

easy cases of knowledge acquisition from the hard ones. One is that the hard cases often involve conceptual construction and conceptual change.

1.1. Knowledge enrichment vs. conceptual construction

Concepts are the atoms of beliefs, the units from which representations of propositions are formed. Most knowledge acquisition consists of acquiring new beliefs (adding new propositions to one's stored knowledge) stated over concepts one already has, or concepts one can construct upon first encountering an exemplar of them. Carey (2009) calls this process "knowledge enrichment." A preschooler's forming a concept of lady bugs, or learning what they are called, or learning what they eat, requires only that they encounter a lady bug, hear it labeled, or be told this fact, because they already have the concepts of animals, animal kinds, and their characteristic diets (see Carey, 2015). Of course, not all cases of knowledge enrichment are easy. Memorizing the names of state capitals or the names for every bone in the body, where the items interfere with each other, is one type of knowledge enrichment that is not easy. So too is straightforward hypothesis testing that is not highly constrained. Finally, so too might be cases of learning procedural knowledge – like the count routine formulated over a long numeral list. Nevertheless, although not all cases of knowledge enrichment are easy, the easy cases of concept learning and knowledge acquisition are *always* knowledge enrichment.

In contrast, sometimes new information presented to children is stated in terms of language that expresses concepts that are incommensurable with those in terms of which the child thinks about the entities referred to. That is, not only do children lack the concepts, they do not yet have the conceptual resources even to express them. For example, telling a 6-year-old child "Gold is an element with atomic number 79," cannot lead to new knowledge shared with a chemist, for the child does not have the concepts *element* or *atomic number* that were first constructed in the 19th century. To understand this sentence, conceptual change is required. Such conceptual changes often unfold over centuries of cultural evolution, always unfold in ontogenetic development only upon months or years of exposure, and often are not achieved by many humans despite years of explicit tutorial in school (c.f. Carey, 2009). Episodes of conceptual change are *never* easy.

Both fast-mapped knowledge enrichment and conceptual change have been extensively explored, but rarely side by side in the same study. Furthermore, very little is known about the domain-general cognitive resources that support these two types of learning. The present study addresses these gaps. We begin by characterizing the two types of learning in general terms and summarize some of what is known about the learning mechanisms that support each. Next, we describe a specific case of conceptual construction, children's acquisition of vitalist biology, which is the target of the present study. We contrast the process of constructing a new framework theory requiring conceptual changes within individual concepts (one kind of conceptual construction, but not the only kind; see Carey, 2009, 2015) with the process of knowledge enrichment where the knowledge is fast mapped. Finally, the present study begins to explore the role of domain-general cognitive resources, focusing here on executive functions and receptive vocabulary, in the two types of learning (conceptual construction and knowledge enrichment).

An episode of conceptual change is characterized by two successive systems of concepts for a common domain of phenomena, Conceptual System 1 and Conceptual System 2 (CS1-CS2). Conceptual systems are characterized through content analyses of the concepts that underlie inferences, patterns of judgments, and explicit explanations and justifications for these inferences and judgments. Establishing conceptual construction requires specifying the ways in which CS2 is qualitatively different from CS1. Conceptual change, one kind of conceptual construction, is change at the level of individual concepts, and consists of differentiations, such that the concept that is undifferentiated in CS1 is incoherent in the light of CS2; coalescences, such that the coalesced concept in CS2 unites entities that cross ontological boundaries in the light of CS1. From the point of view of essentialist analyses of conceptual structure (Ahn & Kim, 2000; Gelman, 2005; Strevens, 2000), conceptual changes also sometimes involve changes in representations of the causally deepest properties of the entities that fall under the concept (Carey, 1985a, 1985b, 2009). Episodes of conceptual change always result in new conceptual primitives, new atoms of beliefs (Carey, 2015).

Carey (2009) characterizes a bootstrapping process ("Quinian bootstrapping"), implicated in at least some episodes of conceptual construction. The essence of Quinian bootstrapping is that the core concepts of CS2 are co-constructed, with their meanings initially almost fully determined by their place in a placeholder structure that represents directly some of their relations with each other, *before* these are deployed in explanation of real world phenomena. In this initial stage of acquisition, the concepts that articulate a placeholder are uninterpreted, or only partially interpreted, with respect to the concepts that articulate CS2. For example, beginning physics students represent the propositions, stated in English and in mathematical equations, "force equals mass times acceleration," or "kinetic energy equals one-half mass times velocity squared" before they have any of the Newtonian concepts *mass, instantaneous velocity, or kinetic energy* (Block, 1986). Filling in the placeholder structures involves extended conceptual modeling (Carey, 2009; Gentner, 2002; Nersessian, 1992), deploying structure mapping, limiting case analyses, thought experimentation, and inductive inference. Consequently, the construction of the concepts of CS2 requires years of instruction, and in the case of these Newtonian concepts, many students never succeed despite such instruction (e.g., Clement, 1982; McCloskey, 1983; McCloskey, Caramazza, & Green, 1980).

Knowledge enrichment, in contrast, is subserved by learning mechanisms that involve hypothesis testing, or statistical learning, over concepts already represented. Examples include associative learning (Rescorla & Wagner, 1972) and some kinds of Bayesian learning, those that involve computations involving hypotheses and priors already represented, as a function of data that is already encoded in terms of the relevant concepts. Sometimes much data is required, so the learning process may be protracted. In other cases, when the learning is very highly constrained, the associative learning or Bayesian hypothesis testing is essentially 1-trial learning. The latter are the easy cases of conceptual development. Episodes of fast-mapped knowledge enrichment differ both quantitatively and qualitatively from those requiring conceptual construction. Carey (1978) coined the term "fast mapping" for the process of acquiring single new words, when the child already has the superordinate for the kind concept the word lexicalizes, and Carey and Bartlett (1978) first demonstrated fast mapping by very young children (2-year-olds). Fast mapping is both *fast* (i.e. the word needs to be heard only once or twice when the referent is clear) and results in a durable mapping (i.e., the child remembers the word weeks or months later, and retrieves it in contexts quite different from that in which it was first encountered). Lexical learning involves associative mappings between newly learned word forms (the encoding of which into the mental lexicon held in long-term memory is constrained by rich phonological, morphological, and syntactic knowledge) and long-term memory representations of the concepts the new words express (subject to constraints on word meanings that are derived from knowledge of syntactic subcategory and knowledge of natural language semantics). No bootstrapping is required for the child to learn the word "hammer" or "goose" or "blue," so long as the child already has the concepts *kind of tool, kind of animal, color.* Rather, the process is one of constrained associative mapping, akin to the Garcia effect (see Gallistel, Brown, Carey, Gelman, & Keil, 1991, for a description of constrained associative learning). Still, there is room for individual differences in the efficiency of lexical learning, differences in the efficiency of encoding the new word form or new concept, the efficiency in forming the mapping between the two in long-term memory, and in the efficiency of retrieval when the context calls for it (Ford & Keating, 1981).

Markson and Bloom (1997) showed that facts like "my grandmother gave me this purse" are also fast mapped, both by preschool children and adults. Learning such facts also involves no bootstrapping. This fact is computed on the fly when the child hears it, for the child has the concepts *grandmother, me,* and *purse*, and comprehends language in terms of its syntactic and semantic combinatorial rules. The mechanisms involved in the acquisition of such facts are also ones involved in the encoding, storing, and retrieving of long-term semantic knowledge. In addition to learning about idiosyncratic facts about particular entities, learning new generic facts about kinds of things (e.g., Octopus blood is blue), also requires no bootstrapping, *if* the child already has the component concepts (in this example, the concepts *octopus, blood, and blue*), and the semantic and syntactic knowledge to compose a representation of this new fact upon first hearing a sentence that expresses it, a representation that then can be stored in long-term memory. One goal of the present study is to explore whether such generic facts are also fast mapped, and whether individual differences in the capacity to learn such generic facts pattern with other measures of individual variation in knowledge enrichment processes, such as measures of receptive vocabulary.

1.2. Acquiring a vitalist theory

Vitalist biology is a framework theory (Wellman & Gelman, 1992) that is shared by adult biologists and non-biologists alike, at least in industrialized cultures (Inagaki & Hatano, 2002; Tardiff, Bascandziev, Sandor, Carey, & Zaitchik, 2017). According to vitalist biology, living organisms are systems that use vital substances such as food, air, and water to produce energy and sustain life. Different body parts have specialized and interrelated functions in this process; they work together in the service of sustaining life via the harnessing energy from the outside world. The failure to do so leads to death, which is the end of life. Crucially, it is the existence of metabolic processes that support the use of vital energy obtained from the outside world that determines whether an entity is a living thing or not. Like all framework theories, a vitalist biology supports many different specific theories consistent with its tenets. For example, sometimes, but not always, "life force" is a theoretical construct that is named (e.g., the "Chi" in East Asian cultures). Death is conceptualized in part in terms of the life force leaving a body. Similarly, the functions attributed to particular body parts vary greatly across different specific vitalist biologies. And of course, professional biologists work with much richer specific theories (e.g., natural selection, the reduction of biological phenomena to chemistry and physics). But these theories do not conflict with the tenets of vitalist biology. Vitalist biology is the ontogenetically earliest conceptual system in which the biological concepts *alive, dead, bodily function, health,* and many others are the same as those represented by adults[1] (Carey, 1985b; Hatano & Inagaki, 1994).

Living things, including both animals and plants, are the entities in the domain of vitalist biology. Young urban preschoolers, who have not yet begun constructing a theory of vitalism, conceptualize animals within a different, non-biological, causal-explanatory system at the core of their essentialized (Gelman, Coley, & Gottfried, 1994) concept *animal.* Supported by very early concepts of causal and intentional agency, manifest even in infancy (Gergely & Csibra, 2003; Saxe & Carey, 2006; Setoh, Wu, Baillargeon, & Gelman, 2013; Woodward, 1998; see Carey, 2009, for review), young preschoolers conceptualize animals as causal/intentional agents, capable of self-propelled motion and goal directed action. Upon learning the placeholder fact that "animals are alive" (it is a placeholder, with respect to CS2, because preschoolers do not yet have the vitalist concept *life*), they attempt to assimilate the concept of being alive to their framework theory of agency. This is the likely reason why young preschoolers often attribute life to the sun and the wind, but deny it to plants, the phenomenon of Piagetian animacy. Moreover, not only is the word 'alive' mapped to the concepts *animate* and *agent*, it is also mapped to *real* and *existent,* because children also learn facts such that dead people (i.e., not alive people) do not exist anymore, or that ghosts are not alive because there are no ghosts. That is, at least among WEIRD (Western Educated

[1] Inagaki and Hatano have used the term "vitalistic causality" to refer to an intermediate explanatory system that falls between intentional causality and mechanical causality. Intentional causality uses the person's intentions to explain biological phenomena and mechanical causality uses physiological mechanisms to explain the same biological phenomena. On their view, vitalistic causality falls in between these two, because it explains biological phenomena by appealing to the activity of internal organs, which to put it in Hatano and Inagaki's (1994, p. 176) words, have "like a living thing, "agency" (i.e., a tendency to initiate and sustain behaviors.)" The bodily activities involve transfer and exchange of a "vital force." The characterization of vitalist biology in the present paper, however, includes both Inagaki's and Hatano's vitalistic and mechanical causalities; both versions of the theories, which Inagaki and Hatano provide evidence for are constructed over the years of 5 to 10 and they fall under the characterization of the framework theory of vitalism as we have characterized it.

Industrialized Rich Democratic) populations (Henrich, Heine, & Norenzayan, 2010), the word 'alive' is mapped to an undifferentiated concept, namely *animate/agent/active/existent/real.*[2] This is the likely reason why when young preschoolers judge an inanimate object like a lamp to be alive, they sometimes justify this judgment by saying that lamps are alive because you can see them or because they have a lamp in their room (Carey, 1985b). The concept *dead* is the end of life in CS2 and it is the negation of the undifferentiated *animate/agent/active/existent/real* concept in CS1. That is, whereas in CS2 *dead* means the end of life resulting from the breakdown of the vital biological processes, in CS1 it means inanimate/inactive/living underground under altered circumstances/absent/non-existent/invisible/imaginary. Furthermore, whereas animals and plants belong to two different ontological categories in CS1 (for plants are causally inert and inactive), they are coalesced into a single category *living thing* in CS2. Finally, whereas the causally deepest properties of animals in CS1 are related to activity, the causally deepest properties of animals and plants in CS2 derive from causal explanatory principles that underlie a vitalist theory of bodily function (Carey, 1985b, 1999; Hatano & Inagaki, 1994; Inagaki & Hatano, 1993, 2002; Slaughter, Jaakkola, & Carey, 1999).

The evidence for this characterization of some of the conceptual changes involved in the construction of a vitalist biology comes from a large body of work, (see Carey, 1985b; Hatano & Inagaki, 1994, for early syntheses), including data from three distinct clinical interviews that investigate the co-construction of quite different vitalist concepts (Slaughter & Lyons, 2003; Slaughter et al., 1999; Zaitchik, Iqbal, & Carey, 2013; Zaitchik, Tardiff, Bascandziev, & Carey, submitted for publication). The first task in this Vitalism assessment battery is derived from Piaget's classic work on childhood animism, focusing on the concept *life* and the distinction between animate and inanimate entities; the second, focuses on the distinction between living and dead entities, the causes and consequences of death (Carey, 1985b; Slaughter & Lyons, 2003). The third builds on the work on understanding of bodily function from the literature on health education (Contento, 1981).

Studies of the acquisition of vitalist biology provide evidence for the differences between knowledge enrichment and conceptual change sketched above. Consistent with the claim that episodes of conceptual change differ quantitatively from episodes of fast-mapped knowledge enrichment is the finding that the concepts *life, death,* and *bodily machinery* undergo the transition from typical preschool, non-vitalist, responses at ages 4, to beginnings of the construction of vitalist theory by some 5-year-olds, to adult-like vitalist responses by ages 10 to 12. The process takes 7 or 8 years; the words "alive" and "dead" are not fast mapped onto the CS2 concepts they express. The much replicated finding that the individual interviews that tap into the concepts *life, death,* and *bodily machinery* are intercorrelated (Slaughter et al., 1999) even after controlling for age, factual knowledge, receptive vocabulary, and executive function (Zaitchik et al., 2013, submitted for publication), supports the claim that conceptual change requires co-construction of a suite of inter-defined new concepts, as required by Quinian bootstrapping.

Johnson and Carey (1998), provided another piece of evidence that knowledge enrichment is qualitatively different from conceptual change. Their study disassociated the two processes by comparing *adults* with Williams syndrome, who have a somewhat preserved ability to learn new facts despite their intellectual disability and their impaired executive functions (Osorio et al., 2012; Porter, Coltheart, & Langdon, 2007; Rhodes, Riby, Park, Fraser, & Campbell, 2010), with a group of *young 6-year-olds* with average receptive vocabulary scores of 100, who have barely begun the construction of the vitalist theory, and with a group of *older children* matched to each Williams syndrome adult on the basis of receptive vocabulary scores (average age, 9;10). The three groups (young children, older children, Williams Syndrome adults) were given tasks that assessed their understanding of vitalist biology, on the one hand, and their accumulated factual knowledge about animals on the other. This is the only study we are aware of that attempted to compare conceptual construction and knowledge enrichment within the same domain, and within the same participants. The Williams syndrome participants performed like the older controls on measures that tapped into knowledge enrichment (factual knowledge about animals), and they performed like the younger controls on measures that tapped into conceptual construction (measures of mastery of vitalist theory). This result suggests that the domain-general cognitive resources that support vocabulary acquisition (relatively preserved in Williams Syndrome) support knowledge enrichment more generally; this hypothesis is directly tested in the present study. Furthermore, this result suggests that the knowledge enrichment mechanisms cannot, by themselves, result in conceptual change.

But what are the domain-general cognitive resources that support conceptual construction? The Williams syndrome participants were matched to older controls on receptive vocabulary only. This means that any number of the remaining domain-general cognitive resources that are impaired in Williams syndrome (e.g. specific EFs, fluid IQ) might be those necessary for conceptual change.

1.3. Domain-general resources supporting the acquisition and deployment of knowledge

Two recent studies have found that measures of EF, specifically inhibition/exogenously cued set shifting and endogenous set shifting, predict performance on the Vitalism battery of tasks that diagnose children's progress toward constructing a Vitalist biology (Zaitchik et al., 2013, submitted for publication). This relationship holds even after controlling for Age, Receptive Vocabulary, Factual Knowledge, Working Memory, and Fluid IQ. Importantly, neither Working Memory nor Fluid IQ predicted performance on the Vitalism battery (Zaitchik et al., submitted for publication). One limitation of these findings, however, is that they cannot differentiate the possibility that EFs are needed *only* for the online *expression* of vitalistic concepts from the possibility that they are

[2] Although cultural context and language shape the specifics of the development of intuitive biology (e.g., see Taverna, Waxman, Medin, Moscoloni, and Peralta (2014), Taverna, Waxman, Medin, and Peralta (2012) and Waxman, Medin, and Ross (2007)), to our knowledge, there is no convincing evidence for any culture in which the word "alive" or "living thing" has commensurable meanings for adults and very young children. In addition, see Astuti, Solomon, and Carey (2004) and Astuti and Harris (2008) for evidence that conceptual change is required for the mastery of adult systems of biology even in non-Westernized, non-urban cultures.

also needed for the *construction* of the vitalist theory. There is no doubt that some of the tasks in the Vitalism battery draw on EF in their online execution (Goldberg & Thompson-Schill, 2009; Shtulman & Valcarcel, 2012; Tardiff et al., 2017). Here we address this interpretative problem through a training study. If EFs are needed for the construction of a vitalist theory, then variance in EFs should predict the degree of pre-training/post-training progress on the Vitalism battery resulting from a training intervention. Furthermore, although two of the tasks in the Vitalism battery (i.e., the Animism and Death interviews) elicit competing incorrect answers that require inhibition (e.g., Harris & Gimenéz, 2005; Shtulman & Valcarcel, 2012; Tardiff et al., 2017), the third (the Body Parts interview) does not elicit such prepotent (incorrect) answers (cf. Tardiff et al., 2017). Thus, a result showing that EFs predict improvement on the Body Parts interview would provide further support to the hypothesis that EFs are involved *not only* in the expression of vitalistic concepts, but that they are also involved in the construction of a vitalist theory.

Previous research has explored the role of EFs in the expression and/or construction of conceptual understanding in other domains, most notably preschool children's Theory of Mind (ToM), and also in preschool children's naïve physics. For example, many studies find that variance in measures of EF predicts variance in measures of Theory of Mind (ToM), especially performance on false belief tasks, controlling for age and receptive vocabulary (see, Devine & Highes, 2014, for a meta-analysis). This finding was extended to a phenomenon within naïve physics: the gravity error committed by 2- and 3-year-olds (Baker, Gjersoe, Sibielska-Woch, Leslie, & Hood, 2011; Bascandziev, Powell, Harris, & Carey, 2016). However, unlike the case of vitalist biology that is constructed in early childhood, whether the developmental changes observed in preschoolers' ToM and naïve physics reflect conceptual change is a subject of controversy. Classic work that argued that the creation of a representational theory of mental states is a construction requiring conceptual change (Bartsch & Wellman, 1995; Perner, 1991; Wellman, 1990) has been challenged by evidence for rich, abstract, representations of mental states, including the capacity to predict the future behavior of an agent on the basis of that agent's false beliefs, even in infancy (e.g., Onishi & Baillargeon, 2005; Senju, Southgate, Snape, leonard, & Csibra, 2011). These results have led some to conclude that the representational resources to express a representational ToM are manifest in infancy and immutable throughout development (e.g., Baillargeon, Scott, & He, 2010; Leslie, 1994; Leslie, Friedman, & German, 2004). These researchers suggest that the *sole* role of EFs is to support the *expression* of ToM on the explicit tasks used with preschoolers.

EFs are undoubtedly needed for the expression of preschoolers' theory of mind in the explicit false belief tasks (Leslie & Polizzi, 1998; Powell & Carey, 2017). However, this fact does not preclude the possibility that they are also drawn upon in the *construction* of conceptual understanding in this domain (Carey, Zaitchik, & Bascandziev, 2015; Carlson & Moses, 2001; Moses & Tahiroglu, 2010; Sabbagh, Xu, Carlson, Moses, & Lee, 2006). One way to differentiate between the two hypotheses is by conducting training studies that investigate whether the variance in EFs predicts variance in *learning* from training. Two recent studies, one in the ToM domain (Benson, Sabbagh, Carlson, & Zelazo, 2013) and one on the gravity error (Bascandziev et al., 2016) explored this possibility. Both found that *learning* from training is predicted by variance in EF. However, an important limitation of these studies is that the training was extremely short, the gains immediate, and it did not result in far transfer, thus providing only very weak evidence in support of the hypothesis that the changes that resulted from the training involved conceptual construction.

1.4. Present study

Here we adapted the Benson et al. (2013) and Bascandziev et al. (2016) strategy to explore the role of domain general resources in a clear case of conceptual change in childhood, by employing longer training with expected far transfer. The training targeted the earliest stages of the construction of a vitalist theory of biology, and compared this process to a hypothesized clear case of fast-mapped knowledge enrichment targeting the acquisition of factual knowledge about animals.

The Vitalism battery, consisting of three clinical interviews used in prior research (Carey, 1985b; Slaughter, Jaakola, & Carey, 1999; Tardiff et al., 2017; Zaitchik et al., 2013, submitted for publication) was administered to participants at pre-training and post-training. These interviews target the concepts of life, death, and the bodily machine. Designed specifically for this study, a Fun Facts pre- and post-training interview probes knowledge of 10 obscure generic facts about animal kinds (e.g., "Where are the ears of a cricket?" Answer: "On their legs."). We expected that at pre-training children would have little or no knowledge of these obscure facts.

The training intervention included 2 repetitions of the answers to all 10 of the Fun Facts questions (one repetition on each of two days of training, less than 5 min total). The vitalist biology training (following Slaughter & Lyons, 2003) was a 2 session (total of 40 to 60 min) intervention concerning the role of internal body parts collaborating with respect to the goals of supporting growth, health, and bodily movement. This collaboration includes bringing in vital substances/energy (air, food, water) from outside the body and distributing them throughout the body. Slaughter and Lyons (2003) found, not surprisingly, that a very similar training intervention targeting the role of bodily function in maintaining life improved performance on the Body Parts interview. More importantly, they found far transfer to the Death interview, providing evidence for the co-construction of vitalist biology and the concepts *life* and *death.* Here we put the hypothesis of co-construction to a much stronger test because neither life nor death were mentioned during our training intervention; indeed, *no* content from the Animism or Death interviews was mentioned during training.

Finally, we administered a battery of EF tasks, as well as a measure of receptive vocabulary, to assess the relations between variance in general cognitive resources and variance in learning outcomes as a result of training on the two batteries (the Vitalist battery and the Fun Facts battery).

We expected to replicate previous findings suggesting that conceptual change and fast-mapped knowledge enrichment are indeed qualitatively and quantitatively different aspects of conceptual development. That is, we predicted that performance on the three Vitalism tasks would be inter-correlated, controlling for receptive vocabulary, age, and accumulated factual knowledge, consistent with the co-construction of the central concepts in vitalist biology, and we expected that although the progress from over an hour of training on bodily function would lead to measureable progress in the early stages of constructing a vitalist biology, such progress would be modest. If learning new generic facts is similar to learning idiosyncratic facts and is therefore fast-mapped, then we should see evidence of fast mapping of the Fun Facts, in sharp contrast to the changes within the child's concepts of life, death and the bodily machine.

New to this study is the exploration of the predictors (EFs and RV) of the variance in children's growth trajectories on Vitalism as a function of the training. If the EFs are drawn upon in the learning process underlying conceptual change, we should find that variance in EF measures will predict the variance in *improvement* on the Vitalism battery, and perhaps, not on the Fun Facts battery. The latter finding would suggest that EFs are not needed for *fast-mapped* knowledge enrichment. Finally, based on Zaitchik et al.'s (submitted for publication) finding that Receptive Vocabulary scores are correlated with accumulated generic academic knowledge, we expected that variance in receptive vocabulary would predict the variance in improvement on Fun Facts, and perhaps not so for the Vitalism battery, which would provide a new source of convergent evidence in support of there being a qualitative difference between the two aspects of conceptual development.

2. Method

2.1. Participants

The sample included 82 kindergarteners (38 girls, 44 boys) with an average age $M = 72.37$ months; $SD = 4.94$; range = 53–86 months. Children were recruited from several schools in the Boston Metro Area from a predominantly white, non-Hispanic, middle-class population. All children were tested in a quiet room or a quiet hallway at their school. Parents received a detailed description of the study and only children whose parents signed a consent form participated in the study.

2.2. Procedure

The design of the present study included a total of five assessment and training sessions: a Pre-Training knowledge assessment, two Training sessions, a Post-Training knowledge assessment, and, lastly, a domain-general abilities assessment session. The sessions were approximately one week apart, administered in the order listed above. The Pre- and Post-Training sessions consisted of a battery of assessments of biological knowledge, and the last session consisted of a battery of EF tasks and a receptive vocabulary task. Assessment always involved one on one testing, whereas the two training sessions involved small groups of 4 to 6 children.

2.2.1. Pre-training/post-training sessions: Knowledge outcome variables: the vitalism and fun facts batteries (see Appendix A for full scripts and scoring schemes)

The Vitalism battery consisted of three different interviews administered in the order: (i) Animism, (ii) Death and (iii) Body Parts. The *Animism* interview (Carey, 1985b; Laurendau & Pinard, 1962; Piaget, 1929) assesses children's concept of living things, in contrast to non-living entities, as reflected in the meaning they have assigned to the word "alive." It begins with questions that directly ask what it means to be alive. In addition, children are asked to name some examples of things that are alive and things that are not alive. Next, children are presented with a list of animals, plants, and inanimate objects, and for each item they are asked if it is alive. In the last section of the interview, four items that appeared on the list (an animal, a plant, and two inanimate objects) are revisited by reminding children what they said about the alive/not alive status of each object/entity, and then by asking them how it is that they know that the object/entity is alive/not alive.

The *Death* interview (Carey, 1985b; Slaughter & Lyons, 2003; Slaughter et al., 1999) begins with questions about what it means to die. Next, children are asked what happens to a person's body when they die, and whether the person's physiological and psychological processes cease upon death. At the end of the interview, children are asked about what might cause a person to die. They are asked to explain/justify all of their responses.

The *Body Parts* interview (Carey, 1985b; Contento, 1981; Slaughter & Lyons, 2003; Slaughter et al., 1999) begins with a series of questions about what various organs are for and what would happen if somebody didn't have those organs. The next few questions ask why we eat food and why we breathe air, what happens to the food and air that go inside our body, and whether the brain is involved in various bodily processes.

The Fun Facts battery consists of questions concerning 10 "Fun Facts," designed to assess factual knowledge about animals that does not depend upon the vitalist framework theory. Children are probed for knowledge of 10 little-known facts about animals, such as "Where are crickets' ears?" Answer: "On their legs."

2.2.2. Scoring

All interviews were videotaped and then transcribed for subsequent offline coding. For the Animism interview, points were given for responses that attributed life to people, animals, and plants and responses that appealed to biological and bodily processes, such as the life-cycle, growth, eating, or breathing. Points were subtracted when children attributed life to inanimate objects, when they denied life to animals or plants, or when they failed to differentiate *alive* from *real* or *existing* (e.g., "I know a lamp is alive, because I have one in my room"). Scores could range from −4 to 19.

For the Death interview, points were awarded for answers that referenced cessation of biological processes, breakdown of bodily function, or decomposition, whereas negative points were awarded for failures to differentiate *inanimate* and *imaginary* from *dead* or for believing that a dead person continues to have physiological and psychological processes (e.g., sleeping, peeing, feeling lonely, etc.). The range of possible scores was from −9 to 8.

For the Body Parts interview, children were given credit for demonstrating knowledge about the function of different organs, for relating the function of body organs to other bodily functions, and for referencing biological processes when answering the questions about eating food or breathing air. Scores could range from 0 to 26.

Points were awarded for factually correct answers on the Fun Facts interview. Possible scores could range from 0 to 10.

2.2.3. Training sessions

The training intervention was modeled after Slaughter and Lyons (2003). The focus of these training sessions was to introduce children to the function(s) of various body parts, to the idea that different body parts are part of a larger system, and to some of the bodily processes that work together as a system to support activity, growth, and health. As in Slaughter and Lyons (2003) training intervention, the training made no mention of death. Unlike the Slaughter and Lyons intervention, this intervention also did not include the words ‘alive’ or ‘keep you alive.’ Thus, the training intervention directly taught material probed on the Body Parts interview, but included no material directly probed on the Animism or Death interviews.

The training was designed to foster Quinian bootstrapping (Carey, 2009). First, the explanations of how bodily processes support *each* biological goal had the same structure, allowing analogical structure mapping among them. The structure in common captures the essential tenets of the framework theory of vitalist biology—vital substance/energy is obtained from the outside environment and is used by internal body parts working together in support of biological goals. This is the placeholder structure with respect to the target concepts of life and death. Second, the explanations that differed among the three biological goals – e.g., the role of food in underlying movement (source of energy) and the role of food in underlying growth (source of building material) – were supported by distinct analogies. Furthermore, the training was closely modeled on that of Slaughter and Lyons (2003), which had been shown to be effective and to lead to some transfer. However, the present study cannot address what aspects of the training intervention are responsible for any learning observed. For that we would need to systematically compare different training interventions. We have taken some initial steps in such an investigation (Bascandziev & Carey, in preparation).

Children were randomly assigned to training groups of 5 or 6 members who were taken to a quiet room for the two training sessions by the experimenter (teacher). Appendix B is the full script of the training protocol, specifying the points that were covered in each session. The teacher engaged the children, elaborated their ideas, and corrected misconceptions, but stuck to the content described in Appendix B. The first training session began with questions about what is inside a human body. Children were asked to point on their bodies where the different organs, such as the heart, the lungs and the stomach, are in the body (see Appendix B for full script). Next, they were introduced to a 3-D anatomical model with removable organs and they were asked to guess which organ was which. All children were given a chance to inspect the organs from the model and each child was assigned at least one organ to pay special attention to. Next, the experimenter introduced the importance of food and air and how they support our bodies’ ability to move, grow, and heal wounds and remain healthy. This explanation included an analogy between our bodies needing energy and cars needing gas. The experimenter emphasized that the energy from food and air goes to all parts of the body. We can move our feet, our fingers, heads, shoulders, and other parts of our bodies. In addition, unlike cars, our bodies grow. An analogy between food and growth and blocks and construction was introduced. The experimenter emphasized that all parts of our bodies grow, and it is very important for the food and the air to reach all parts of the body, because if they don’t, they can’t grow, and also, those parts of the body will get very sick. But how is it that the food we eat and the air we breathe reach all parts of our bodies? This question was the focus of the next section in which the experimenter explained how the food gets broken down into small pieces as it moves through the mouth, stomach, and intestines, where it then enters the bloodstream. This led to talking about blood vessels and about how they are used to transport the blood rich with nutrients and air to all parts of the body, just like buses transport passengers around cities. At this point, the function of the heart was introduced as a pump that helps move the blood around the body. The kidneys were also introduced as an organ that helps the blood to stay clean just as a carwash helps cars to stay clean. At the end of the vitalism training, the brain was introduced as the organ that is responsible for our psychological states and that controls the function of all other organs. Still, even though the brain is the boss of our body, it needs food and air, just like any other organ in our body. Finally, the experimenter told children what the answers to the fun fact questions were, while showing them pictures and drawings that illustrated the answers (e.g., a picture of a dolphin sleeping with one eye closed, an ostrich egg, etc.). The training session ended with a question-and-answer period concerning the Bodily Function training. Children were asked if they remembered the function of the organ to which they were assigned, why that organ is important, and where it should be put in the 3-D model.

The exact same content was covered in the second training session, which differed from the first in using a poster with a drawing of the human body and paper cut out organs rather that an anatomical 3-D model. The organs had Velcro on one side, so they could be attached and removed from the poster. Children were assigned a particular organ and were asked to tell the group what they remembered about the function of that organ, and they were asked to attach it where they thought it should go on the poster. In addition, the experimenter repeated all the important information that was covered in session 1. Similarly, for the Fun Facts portion of the training, the experimenter asked children if they remembered what the picture/drawing was about and if they remembered what the question and the answer to that question was. All the relevant information was repeated, regardless of how well children remembered the answers to each question.

2.2.4. Domain general abilities assessment session

Children received three executive function tasks administered in the following order: (i) Hearts and Flowers; (ii) Flanker; and (iii) Verbal Fluency, followed by a test of receptive vocabulary.

The *Hearts and Flowers* task (Davidson, Amso, Anderson, & Diamond, 2006; Diamond, Barnett, Thomas, & Munro, 2007) is a computer task that begins with a simple rule asking children to press the button that's on the same side as the heart. The heart appears either on the left or on the right side of the computer screen. After completing a few practice trials, children receive a block of heart trials. Next, a new rule is introduced asking children to press the button that's on the opposite side of the flower. The flower appears either on the left or on the right side of the screen. Once again, after a few practice trials, children receive a block of flower trials. In the last block, children receive randomly mixed trials where the instruction is that they will play both games at the same time. The last block of trials was of interest in this study, because these trials tap into children's ability to hold both rules in mind (working memory), their ability to inhibit one of the responses in order to press the correct button (inhibitory control), and their ability to flexibly switch between the two rules (set shifting). Children's accuracy was the dependent measure on this task.

On the *Flanker* task (Davidson et al., 2006; Diamond et al., 2007), the stimuli are lines of fish, such that the middle fish is sometimes facing in the same direction as the flanking fish, and sometimes in the opposite direction. The fish are sometimes all pink and sometimes all blue. The first rule: when the fish are blue, the task is to feed the middle fish, by pressing the button where the middle fish is facing (i.e., the left button if the fish is facing left; the right button if the fish is facing right). After a few practice trials, children receive a block of trials with blue fish. On some trials, the middle fish is flanked with other fish that are facing in the same direction (congruent trials), and on some trials, the middle fish is flanked with fish that are facing in the opposite direction (incongruent trials). There are also trials in which there are no flankers (neutral trials). Then, children complete a block of trials following the second rule: when the fish are pink, the task is to feed the fish that are on the flanks. The flanking fish are all by themselves on some trials, they are accompanied by a middle fish that is facing in the same direction, or they are accompanied by a middle fish that is facing in the opposite direction. Finally, in the last block of randomly mixed trials, children are instructed that they will play both games at the same time. When the fish were blue, they need to press where the middle fish is facing, and when the fish were pink, they need to press where the flanking fish is facing. As with Hearts and Flowers, accuracy on the mixed trials was the dependent measure for this task.

The *Verbal Fluency* task consists of two subtests: Animal Naming and Food Naming. In this task, children are asked to name as many animals/foods as they can in just one minute. Success on this task depends on children's ability to use the abstract superordinate concepts *animal* and *food* to search their huge lexical database, and to flexibly shift between different subcategories of animals and foods. This task requires children to find subcategories of animals to list, and to monitor when one category (e.g., farm animals) is exhausted and then switch to a different category (e.g., wild predators), and thus draws on *endogenous* set shifting and data base search, as well as proactive EF processes (Chatham, Frank, & Munakata, 2009). Children's scores reflect the number of animals/foods they named, minus any repetitions or incorrect responses. Children of these ages know the names of dozens of animals and foods (i.e., they can correctly answer questions such as: "is a cheetah an animal, are turtles animals, are potatoes food, is ice cream food…?) At the age of our sample, poor performance consists of naming only 2 or 3 animals; the problem is in the initial search of this huge database, and perhaps the difficulty in inhibiting the first animals or foods thought of to generate further examples.

Receptive Vocabulary. As a measure of factual knowledge, we administered a subtest of WPPSI-III's verbal IQ test in which children are asked to select one of four pictures that corresponds to a word. See Zaitchik et al. (2013, submitted for publication) and Johnson and Carey (1998), for theoretical and empirical rationales for taking the raw score on a receptive vocabulary measurement as an indicator of accumulated factual knowledge.

3. Results

We first present the descriptive statistics of the pre- and post-training performance on the vitalism interviews and the Fun Facts questions and we ask whether children improved between pre- and post-training on these interviews. Then we present the results on the EF tasks. Next, we present a series of regression analyses where the outcome variables are either pre-training or post-training scores on Vitalism and Fun Facts and the predictor variables are EFs and Receptive Vocabulary (RV). Finally, we present mixed model analyses that address the main question of this paper: whether EFs and RV predict who will benefit most from training. In particular, we explore whether EFs are uniquely predictive of variance in progress toward conceptual construction, as assessed by the Vitalism battery, whereas RV is uniquely predictive of variation in knowledge enrichment, as assessed by the Fun Facts battery.

3.1. Biology battery (descriptive statistics and pre-training/post-training improvement)

Children's responses to the three Vitalism interviews, as well as their answers to the Fun Facts questions, were transcribed and coded interview by interview, the coder blind to the identity of the participants and thus to their performance on the other knowledge assessment tasks and the predictor variable tasks. Each interview was coded by two independent coders. Inter-rater agreement, calculated using the intraclass correlation coefficient, was high, ranging between 0.88 and 0.98. Disagreements were resolved via discussion.

3.1.1. Descriptive statistics, computing composite outcome variable

Table 1 presents the descriptive statistics of the raw pre-training (Panel 1) and post-training (Panel 2) scores on the three interviews designed to tap into children's vitalist theory of biology (Animism, Death, and Body Parts) and the Fun Facts interview, designed to tap into obscure factual knowledge about animals. Inspection of Table 1 confirms that most children at this age are in the early stages of constructing a vitalist theory of biology. The mean scores on Animism, Death, and Body Parts are much lower than the possible maximum scores both at pre- and post-training. In addition, the Fun Facts interview indeed quizzed children about obscure facts, as shown by the very low average pre-training score. Importantly, however, not all children performed identically. The

Table 1
Descriptive statistics of the raw pre- and post-training scores on the four biology interviews (n = 82).

	Animism	Death	Body Parts	Fun Facts
Panel 1: Pre-training				
Mean Raw Score	5.98	0.83	9.79	1.95
SD	3.39	2.73	3.61	1.06
Min	0	−6	3	0
Max	16	6	20	4
Possible Range	−4 to 19	−9 to 8	0–26	0–10
Panel 2: Post-training				
Mean Raw Score	7.45	1.27	11.55	7.18
SD	3.71	2.60	4.35	1.96
Min	0	−5	2	1
Max	17	6	21	10
Possible Range	−4 to 19	−9 to 8	0–26	0–10

Table 2
Means of standardized scores of pre- and post-training performance on all interviews and results of paired *t*-tests demonstrating pre-/post-training improvement (both pre- and post-training scores were standardized by using the pre-training means and standard deviations) (n = 82).

	Animism	Death	Body Parts	Fun Facts
Pre-Training	0	0	0	0
Post-Training	0.43	0.16	0.49	4.9
t value	5.06	2.21	4.94	25.83
p value	< .001	< .05	< .001	< .001
95% CI of the difference	[0.26–0.60]	[0.02–0.30]	[0.29–0.68]	[4.53–5.29]

standard deviations and the range of scores that children earned in these interviews show that there was considerable individual variability.

3.2. Pre-post improvement

Panel 2 of Table 1 presents the post-training scores. Table 2 presents the pre-training and the post-training scores standardized relative to the pre-training means and standard deviations, which is why the means on all pre-training measures are 0 in Table 2. Table 2 also presents the results of paired *t*-tests, establishing whether children improved significantly on each measure.[3] Inspection of Tables 1 and 2 reveals that children improved on each of the four separate outcome measures.

Given that we directly taught the answers to the Fun Facts questions at the end of each training session, we expected children would improve from pre-training to post-training. Indeed, they did so, by 5 SDs ($t(81) = 25.83$, $p < .001$). This immense change reflects children's near-floor performance at pre-training on these obscure facts, as well as that merely adding new facts to one's body of knowledge is relatively easy. It extends Markson's and Bloom's finding of fast-mapping of idiosyncratic facts (e.g., my uncle gave me this) to fast-mapping of newly encountered generic facts (e.g., cows have four stomachs).

Just as with Fun Facts, the capacity for straightforward belief revision and learning new facts should lead to gains on the Body Parts interview, and indeed it did by 0.5 SD ($t(81) = 4.94$, $p < .001$). Children received an hour of discussion while engaging with physical representations of the human body (3-D anatomical model, poster with removable organs), learning about how the organs of the body work together to harness the energy and material from vital substances obtained from outside the body (food, air) in support of growth, activity, and healing. The magnitude of this change, however, is one tenth that on the Fun Facts; we return to this difference below (Section 3.8).

In contrast, that the training would impact the Death and Animism interviews (by 0.16 and 0.43 SDs respectively) was far from a foregone conclusion. Nothing in the training taught any facts about death or which entities in the world are alive and why. But despite no facts being provided during the training sessions that could have directly helped children improve on the Animism and the Death interviews, children improved on both (Table 2). That is, even though children did not hear the words "alive," "life," "dead," or "death" during the training session, at post-training they could more correctly classify and explain which entities are alive and they could better articulate what it means to die, what happens to people when they die, and what causes death, in terms of vitalist theory. Although this study was not designed to answer which aspects of the training helped children, this finding confirms that the training as a whole moved children along in their construction of the framework structure of the theory of vitalism; namely, some

[3] Scores were standardized in terms of pre-training means and standard deviations for the purpose of creating composite variables, and to have a standard unit (SDs of pre-training scores) across tasks to compare degree of change on each interview. Of course, the *t*-values and significance levels of the change between pre-training and post-training are identical when the input to the analyses are the raw scores depicted in Table 1 or the standardized scores in Table 2.

Table 3
Bivariate and partial correlations (below diagonal, controlling for age and receptive vocabulary) among the three bio interviews at pre-training (n = 82).

	Animism	Death	Body Parts	Fun Facts
Panel 1. Pre-Training				
Animism	*	0.45^{***}	0.39^{***}	0.29^{**}
Death	0.34^{**}	*	0.25^{*}	$0.19^{\sim}$
Body Parts	0.35^{***}	0.22^{*}	*	0.04
Fun Facts	0.23^{*}	0.09	0.01	*
Panel 2. Post-Training				
Animism	*	0.61^{***}	0.47^{***}	0.40^{***}
Death	0.55^{***}	*	0.50^{***}	0.45^{***}
Body Parts	0.46^{***}	0.49^{***}	*	0.28^{**}
Fun Facts	0.28^{**}	0.35^{**}	0.26^{*}	*

Key:
~ p ≤ .10.
* p ≤ .05.
** p ≤ .01.
*** p ≤ .001.

reorganization of their concepts *alive* and *dead,* including some differentiation of *alive* from *animate, real,* and *existing,* and some differentiation of *dead* from *inanimate, imaginary, invisible,* and from *opposite of alive* to *end of life.*

3.3. Coherency among the vitalism interviews

Given that the vitalism interviews are designed to measure interrelated concepts of vitalist biology, we predicted that they would be correlated both at pre-training and at post-training. As seen in Table 3, this prediction was confirmed, even after partialling out Age and Receptive Vocabulary (partial correlations below diagonal). The coherence of the three biology interviews (Animism, Death, and Body Parts) licensed the creation of a composite outcome variable (Vitalism), separately computed for pre-training interviews and post-training interviews by averaging the standardized scores of the three interviews (both Pre- and Post-training scores were standardized relative to Pre-training means and standard deviations). On theoretical grounds, because the Fun Facts have nothing to do with Vitalist theory, we kept the Fun Facts scores (pre-training and post-training) as separate outcome variables.

Notice, however, that the pre-training Fun Facts scores were correlated with Animism scores, even controlling for Age and Receptive Vocabulary. This was unexpected, because the Fun Facts do not diagnose an understanding of vitalist biology, nor do they concern what "alive" means, and what entities are and are not alive, and also because of the near floor pre-training performance of the group as a whole on these obscure facts. Furthermore, although the Fun Facts task was not designed to be a pre-training measure of accumulated factual knowledge in general or of accumulated factual knowledge about animals in particular, pre-training scores of Fun Facts were correlated with Receptive Vocabulary scores ($r(80) = 0.23$, $p < .05$). If Receptive Vocabulary is, as we assume, a proxy for general factual knowledge (see Zaitchik et al., submitted for publication), then this result suggests that pre-training performance on Fun Facts can also be considered a (very weak) measure of general factual knowledge about animals. Thus, to further confirm that intercorrelations among the three vitalism interviews remain significant after controlling for Age and Factual Knowledge, we asked whether they remain significant after controlling for Age, Receptive Vocabulary, *and Fun Facts as well.* Indeed, they do. The intercorrelations remained virtually unchanged (Animism-Death: $r(77) = 0.32, p < .01$; Animism-Body Parts: $r(77) = 0.36$, $p < .01$; Death-Body Parts: $r(77) = 0.22, p \leq .05$), supporting the conclusion that the correlations among the vitalism interviews does not simply reflect accumulation of factual knowledge.

A comparison of the strength of the intercorrelations at pre-training and post-training is informative. On the one hand, if theory construction is simply a matter of accumulating factual knowledge, then the training would result mainly in changes on the Body Parts interview, because no factual knowledge directly probed on the Death and Animism interviews was imparted during training. Thus, we would expect the three interviews to be less tightly intercorrelated at post-training than at pre-training. On the other hand, if the training succeeds in instilling or advancing the basic framework of a vitalist theory—of vital substances obtained from outside the body, the idea of integrated bodily functions subserving goals such as growth, bodily repair, and the capacity for movement and activity—and this vitalist theory in turn provides a framework for reinterpreting phenomena that are probed on the Animism and Death interviews, then we would expect individual children to improve on all three interviews. We might even see stronger intercorrelations among the three interviews at post-training than at pre-training, because it is by virtue of vitalist biology that the concepts of life, death, and the bodily machine are interrelated. By hypothesis, the stronger intercorrelations would be due to individual differences among children in their progress at constructing a vitalist understanding of bodily function because of the training, and the consequent reanalysis of the concepts probed on the Animism and Death interviews.

Inspection of Table 3 reveals that the magnitude of the correlations at post-training is approximately 1.5 times *greater* than those at pre-training. To test whether this apparent increase in coherency was significant, we performed a one-sample comparison of the pre-training to the post-training Cronbach's alpha coefficient (Diedenhofen & Musch, 2016; Feldt, Woodruff, & Salih, 1987). Given that Cronbach's alpha is a statistical measure of the internal consistency of items that are assumed to measure the same underlying construct, we asked whether the internal consistency of the vitalism battery items across all interviews would increase when the items remain the same, but the underlying construct undergoes a change. The answer to this question is yes. The pre-training Cronbach's alpha was 0.77 and the post-training Cronbach's alpha was 0.84. The difference between the two coefficients is statistically significant ($\chi^2(1) = 8.04$, $p < .01$).[4] Of course, this would have been a trivial result if children were taught facts that could directly affect their performance on all three interviews. Thus, it is very important that the increase in intercorrelations among the items on the vitalism battery was a result of a training that did not include any new facts about Animism or Death. That is, children were not taught facts such as the sun is not alive; plants are alive, dead bodies decompose, and so on. Yet, after receiving the vitalism training, children's scores on all three interviews became more closely intercorrelated. This is particularly strong evidence that one of the factors driving the intercorrelations among the Animism, Death, and Body Parts scores is the degree of progress children have made in constructing the framework vitalist theory of biology.

3.4. The relationship between Fun Facts and Vitalism scores

A final notable result from Table 3 is that post-training performance on the Fun Facts interview was robustly correlated with post-training performance on *each* vitalism interview ($p < .02$, after partialling out Age and Receptive Vocabulary). The facts taught could have no impact on the child's performance on any Vitalism interview. Facts like the location of a cricket's ears are irrelevant to the function of the heart, whether dead people need to pee, and which things in the world are alive. For this reason, learning the fun facts could not have led to increased coherency among the vitalism interviews at post-training compared to pre-training. That this is so was confirmed by an analysis showing that the post-training intercorrelations among the vitalism interviews remained virtually unchanged (Table 3, Panel 2) even after partialling post-training Fun Facts performance in addition to partialling out Age and Receptive Vocabulary (Animism-Death: $r(77) = 0.50$, $p < .001$; Animism-body Parts: $r(77) = 0.41$, $p < .001$; Death- Body Parts: $r(77) = 0.44$, $p < .001$).

How might we explain the correlations between the post-training Fun Facts scores and the post-training Vitalism interview scores? To interpret this finding, it is important to note that the post-training variables reflect the variance that was present at pre-training and the added variance contributed by the training itself. The Fun Facts variable had relatively little variance at pre-training, because of the obscurity of the facts. All children were exposed, equally, to the answers of the Fun Facts questions during training. Thus, the added variance during training makes the post-training Fun Facts variable a more representative measure of children's *ability to learn new facts* than is the pre-training variable, which also incorporates variance in children's encounters with these facts. That post-training Fun Facts scores predict Vitalism post-test scores suggests that fast-mapped generic fact learning, supported by knowledge enrichment mechanisms, is *important* to conceptual construction, even in the face of all the evidence reviewed above that knowledge enrichment mechanisms alone are not *sufficient* for conceptual construction.

Clearly, learning new facts is an important part of the process of conceptual change. It stands to reason that the capacity to learn new facts might be predictive of the capacity for theory construction, simply because the process of theory construction involves learning new facts. Indeed, the vitalism training intervention teaches the child many facts about the human body, presented in the context of a vitalist understanding of bodily function. Moreover, if the child did not understand these facts, the child would be unlikely to understand the tenets of vitalist theory that were taught, the structure of which was reinforced with many analogies. Nonetheless, the analyses presented above (i.e., the far transfer on the Death and Animism interviews; the increased coherence among the vitalism interviews at post-training compared to pre-training) show that theory construction is not *merely* the consequence of belief revision and accumulating new factual knowledge.

If the processes underlying the performance and improvement on the vitalism battery are, at least in part, different from the processes underlying the performance and improvement on Fun Facts, we should expect that different predictor variables will be related to the two underlying processes. That is, the variable(s) that predict individual differences in the ability to construct a theory of vitalism (reflected in pre-training, post-training scores, and the amount of change on the Vitalism battery due to the training) should be different than the variable(s) that predict accumulated factual knowledge and the capacity to learn and retain new factual knowledge (pre-training scores, post-training scores, and the amount of change on the Fun Facts due to the training). Especially important will be analyses of the predictors of amount of change as a function of training (post-training gains, relative to pre-training scores) of the composite Vitalism scores and Fun Facts scores.

3.5. Descriptive statistics, creating composite predictor variables (EF and Receptive Vocabulary)

We first present the relations among the predictor variables that justify the construction of composite predictor measures. For this analysis, simple correlations are the relevant data. We describe the performance on each of the four EF tasks and Receptive Vocabulary (Table 4, Panel 1) and then present the correlations among these measures (Table 4, Panel 2). The proportion correct scores (Panel 1) on Hearts and Flowers and Flanker were similar, as were the mean scores on the two verbal fluency tasks (Foods and Animals). In addition, the measures of variability show that children's scores varied widely on all four EF tasks and on the Receptive Vocabulary task.

The correlations presented in Panel 2 show that, as anticipated, the two verbal fluency tasks were correlated with each other.

[4] We also computed Cronbach's Alpha coefficients by treating the composite scores on the three biology interviews as items (i.e., a 3-item biology test). The results were consistent with the analysis above. Cronbach's Alpha was .63 at pre-training and it was .76 at post-training. The difference between the two coefficients is statistically significant ($\chi^2(1) = 6.16$, $p = .01$).

Table 4
Descriptive statistics of the four EF tasks and Receptive Vocabulary (Panel 1) and correlations among those variables (Panel 2) (n = 82).

	Hearts & Flowers	Flanker	VF Foods	VF Animals	Receptive Vocabulary
Panel 1: Descriptive Statistics (Hearts & Flowers and Flanker: proportion correct; Verbal Fluency and Receptive Vocabulary: raw scores)					
Mean	0.81	0.83	12.32	13.17	31.07
SD	0.15	0.15	4.68	4.21	3.17
Min	0.41	0.30	0	1	22
Max	1	0.98	24	23	36
Panel 2: Bivariate correlations among the four EF tasks and receptive vocabulary					
Hearts & Flowers	*	0.12	0.02	0.04	0.34**
Flanker		*	0.23*	0.24*	0.21~
VF Foods			*	0.32**	0.17
VF Animals				*	0.17
Receptive Vocabulary					*

Key: ***p ≤ .001.
~ p ≤ .10.
* p ≤ .05.
** p ≤ .01.

Furthermore, the bivariate correlations between the Flanker task and the two verbal fluency tasks were also significant. This result is interpretable because both the Flanker and the Verbal Fluency tasks involve set shifting, although exogenous (cued) set shifting for the former and endogenous set shifting for the latter. Surprisingly, however, the correlations between Hearts and Flowers and the other EF tasks were not significant. Rather Hearts and Flowers was correlated only with Receptive Vocabulary. The lack of significant relationship between Hearts and Flowers and the remaining three EF tasks and the stable relationship with Receptive Vocabulary suggest that the major source of variance on the Hearts and Flowers task was not related to EF, but was related to lexical knowledge. One possible interpretation of this unexpected result is that there might have been variability in children's ability to understand the phrase "opposite side" (used in the HF task) to guide their behavior. Indeed, the second block that asked children to press the button on the "opposite side" from the flower was also correlated with Receptive Vocabulary, controlling for Age ($r(79) = 0.40, p < .001$). Given that Hearts and Flowers was not correlated with any of the remaining EF tasks, we dropped this variable from further analyses and we computed a composite EF variable by standardizing and averaging the two verbal fluency tasks and the Flanker task. The variable Age was centered with a mean of 0.

On theoretical grounds, we also kept raw scores on Receptive Vocabulary as a separate measure of children's accumulated factual knowledge. Receptive Vocabulary is often taken as a measure of Verbal IQ (crystallized intelligence), i.e., as a measure of individual variation in the capacity to *learn* new factual information (e.g., the names for objects). In this use, the scaled scores (adjusted for age) are the relevant variable, but due to our restricted age range, the correlations between raw and scaled scores was high ($r(80) = 0.94$, $p < .001$) and thus we use the raw scores alone both as a measure of individual variation in accumulated factual knowledge and as a measure of variation in ability to learn new factual knowledge. Our sample had a mean Receptive Vocabulary scaled score of 12.79, which is 0.93 standard deviations above the national mean, not unexpected given our sample's demographics.

The two knowledge outcome variables were designed to measure two different kinds of knowledge of biology. The composite Vitalism outcome variable reflects children's understanding of vitalism, which includes a vast body of factual knowledge, organized into a theoretical structure connecting the different facts into a coherent causal-explanatory framework that gives meaning to the concepts embedded in it, including the concepts of *life* and *death* that are co-constructed with the construction of the theory of vitalism. The Fun Facts outcome variable was designed to measure children's ability to learn new facts that are not related to each other in any meaningful way, except that they are all about animals, and that they are facts formulated in terms of concepts the child already possesses. The remaining analyses concern what domain general variables, if any, predict these two outcome variables, both pre-training and post-training, and importantly, which variables, if any, predict which children benefit more (i.e., variability in change from pre- to post-training) on the composite Vitalism measure and on the Fun Facts measure.

3.6. Predicting Vitalism and Fun Facts pre-training scores

3.6.1. Bivariate correlations

Replicating previous findings, we found that Vitalism-Pre was positively correlated with EF ($r(80) = 0.24, p < .05$), with RV ($r(80) = 0.40, p < .001$), and Age ($r(80) = 0.28, p < .05$) (see Zaitchik et al., 2013, submitted for publication). That is, children who scored higher on EF, RV, and who were older also scored higher on Vitalism at pre-training. Similarly, Fun Facts performance at pre-training was correlated with EF ($r(80) = 0.22, p < .05$) and RV ($r(80) = 0.23, p < .05$). That is, children who scored higher on RV and EF also scored higher on Fun Facts at pre-training.

3.6.2. Predicting Vitalism-Pre

Table 5, Panel 1 presents the regression coefficients of a regression analysis examining the effects of all three predictor variables (EF, RV, and Age) in the same model on the outcome variable Vitalism-Pre. The intercept represents the predicted performance on the outcome variable when all other variables included in the model are 0 (recall that because all variables were standardized, 0

Table 5
Regression coefficients from multiple regression analysis with pre-training scores (Panel 1) and post-training scores (Panel 2) as outcome variables.

Predictors	Pre-Training Outcome Variables	
	Vitalism-Pre Beta coefficients	Fun Facts-Pre Beta coefficients
Panel 1. Multiple regression analysis of the pre-training performance on the two outcome variables		
Intercept	0	0
Age	0.03~	−0.001
Receptive Vocabulary	0.27***	0.19~
Executive Function	0.06	0.18

Predictors	Post-Training Outcome Variables	
	Vitalism-Post (Beta coefficients)	Fun Facts-Post (Beta coefficients)
Panel 2. Multiple regression analysis of the post-training performance on the two outcome variables		
Intercept	0.36***	4.9***
Age	0.006	−0.02
Receptive Vocabulary	0.26**	0.76***
Executive Function	0.21*	0.18

Key:
~ $p \le .10$.
* $p \le .05$.
** $p \le .01$.
*** $p \le .001$.

represents average performance). We found that Age trends toward being a significant predictor, when RV and EF are controlled for; RV is a significant predictor, when Age and EF are controlled for, whereas EF is not a significant predictor. Controlling for Age and EF, the predicted score on the composite Vitalism at pre-training is 0.27 higher for every 1-unit difference in RV. Thus, variance in RV predicts variance in Vitalism-Pre, over and above the contribution of EF and Age.

The lack of a significant relationship between Vitalism-Pre and the composite EF variable is inconsistent with previous findings from two different samples of approximately 80 children each. Both studies found a positive relationship between Vitalism scores and EF even after controlling for Age and factual knowledge (Zaitchik et al., 2013, submitted for publication). Very likely, this inconsistency is due to the fact that the children in the present study were on average six months younger than those in Zaitchik et al.'s (2013, submitted for publication) studies. The children in the present study were probably at the very beginning of the process of constructing a theory of vitalism—such that there was little variance in the progress of the relevant conceptual construction—leaving accumulated factual knowledge, and perhaps variability in the efficiency of learning new factual knowledge, as the major source of variance at pre-training. For this reason, it is important to investigate the relationship between the predictor variables and post-training performance (see Section 3.7), where an important source of individual differences at post-training will be due to children's ability to advance their understanding of vitalism.

3.6.3. Predicting Fun Facts-Pre

Table 5, Panel 1 presents the regression coefficients of a multiple regression analysis with Fun Facts-Pre as the dependent variable, and EF, RV and Age entered into the model as predictors. Age was not a significant predictor, controlling for RV and EF. Controlling for Age and EF, RV was trending toward being significant, and controlling for Age and RV, EF was not a significant predictor. Thus, neither EF nor RV predict variance in Fun Facts-Pre over and above the other predictor and Age. It is important to note, however, that children were near floor on this task at pre-training. Therefore, it is important to examine the relationship between the predictor variables and *post-training* performance on Fun Facts, since all children heard the same facts equally during training, but some learned more than others.

3.7. Predicting Vitalism and Fun Facts post-training scores

Next, we examined the relationship between the predictor variables and the outcome variables at post-training. The post-training variables include variance that children brought to pre-training and new variance that was introduced by the training itself.

3.7.1. Predicting Vitalism-Post

Table 5, Panel 2, presents the regression coefficients of a multiple regression analysis with Vitalism-Post as an outcome variable and Age, RV, and EF as predictor variables. The intercept, which is significantly different from 0, represents the predicted performance on Vitalism-Post for a child who scored 0 on all other variables (i.e., an average child). Because the intercept represents the performance of an average child at post-training, the intercept values are equivalent to the numbers reported in Table 2 above (note that 0.36 is the average performance at post-training across the three interviews (Animism, Death, and Body Parts), and 4.9 is the average post-training performance on Fun Facts). Furthermore, because both the pre- and the post-training scores were standardized relative to the pre-training means and standard deviations, the average performance at post-training also represents the average increase between pre- and post-training. As seen in Table 5, Age was not a significant predictor when RV and EF were controlled for, but both RV and EF were significant predictors. Controlling for Age and EF, for every 1-unit difference in RV, children on average score 0.26 higher on Vitalism-Post. Similarly, controlling for Age and RV, the predicted score on Vitalism-Post is 0.21 higher for every 1-unit difference in EF. This finding is consistent with previous findings, showing that variance on measures of RV and EF each predict unique variance in children's understanding of vitalism (Zaitchik et al., 2013, submitted for publication).

3.7.2. Predicting Fun Facts-Post

Table 5, Panel 2, also presents the results of a multiple regression with Fun Facts at post-training as an outcome variable and Age, RV, and EF as predictor variables. The intercept, which is significantly different from 0, represents the predicted score on Fun Facts-Post when all predictor variables are 0. As seen in Table 5, neither Age nor EF were significant predictors of Fun Facts-Post when the other variables were controlled, whereas RV was. Controlling for Age and EF, the predicted Fun Facts Post-training score is 0.76 higher for every 1-unit difference in RV. Thus, the post-training variance on the outcome variable that was designed to measure children's ability to learn new facts was overwhelmingly predicted by children's RV scores, over and above Age and EF.

In sum, both post-training outcome variables are related to Receptive Vocabulary and only the Vitalism outcome variable is related to Executive Function. This finding is consistent with the hypothesis that EF is needed for the process of constructing new theories and new conceptual knowledge. However, the post-training variables include variance contributed by the training itself and variance carried over from pre-training. The strongest test of the hypothesis that EF is uniquely drawn upon in theory construction and conceptual change would be an analysis that asks the following question: controlling for pre-training performance and thus isolating the variance introduced by the training itself, who were the children who benefitted the most and who were the children who benefitted least from training?

3.8. Predicting the rate of change[5]

The primary goal of the present experiment is to explore the hypothesis that different variables might predict *improvement* as a consequence of training on our two outcome variables (Vitalism and Fun Facts), and in particular, participants higher in executive function may make more progress in the construction of a vitalist theory than those lower in executive function. To test these hypotheses, we fitted mixed models to data for each of the two outcome variables (Vitalism and Fun Facts) with Age, Receptive Vocabulary, and Executive Function as predictor variables (Bascandziev et al., 2016; Singer & Willett, 2003). Table 6 presents the *rate of change* coefficients.[6] The intercept here represents the average increase from pre- to post-training for a child who scored 0 on all predictor variables included in the model (i.e., average EF, RV, and Age). As mentioned above, because we standardized the pre-training and the post-training scores by using the pre-training means and standard deviations, the average pre-training performance is 0 and the average post-training performance is the same number as the average increase from pre-training to post-training. That is, 0.36 is the average post-training performance on Vitalism, the average *increase* (from 0 at pre-training to 0.36 at post-training), and the average slope (rate of change). Similarly, 4.9 is the average post-training performance on Fun Facts, the average *increase* (0–4.9) from pre-training to post-training on Fun Facts, and the average slope (rate of change). Hence, we use the locutions "rate of change" and "amount of improvement" interchangeably. Age predicts amount of improvement only when the outcome variable is Vitalism. This small but statistically significant effect means that controlling for RV and EF, for every 1-unit difference in Age, there is a 0.03 *decrease* in the rate of change on the Vitalism composite scores. Younger children have more room to improve, both with respect to absorbing new factual knowledge and constructing the theory of vitalism. On the other hand, the very small parameter value is consistent with the possibility that this effect reflects noise.

Turning to the measures of domain-general resources (RV and EF), we see that scores on Receptive Vocabulary were not significantly associated with the rate of change in Vitalism, but were with the rate of change in Fun Facts. Controlling for Age and EF, for every 1-unit difference in RV, on average, there is a 0.57 increase in the rate of change of Fun Facts. The Executive Function predictor shows the opposite pattern across the models for the two outcome variables. Namely, scores on EF significantly predicted the rate of change of Vitalism, controlling for Age and RV, but were unrelated to the rate of change of Fun Facts. Controlling for Age and RV, for

[5] Here, we use the term 'rate of change' to refer to 'improvement' because we used mixed models to analyze whether EFs and RV predict the variance in children's individual rates of change from pre-training to post-training. That is, some children's growth trajectories on Vitalism and Fun Facts have positive slopes, of varying steepness, some are flat, and some may have a negative slope as a result of receiving the training intervention. The mixed model analysis presented here asks whether there is an interaction between our predictor variables (EFs and RV) and Training, where training can take only two values: Pre-Training = 0 and Post-Training = 1. A significant interaction would therefore mean that the slopes (from pre-training to post-training) differ according to levels of EF (or RV) (see Bascandziev et al., 2016 for a mathematical description of the model). For example, if children's growth trajectories from pre- to post-training do not systematically differ as a function of the varying EF scores, then the EF × training interaction will not be significant. Conversely, if children's slopes from pre- to post-training differ systematically as a function of EF (or RV), then there will be a significant interaction and therefore, the predictor EF (or RV) will be a significant predictor of change (improvement) over time. It is important to note, however, that because there are only two points in time (pre-training and post-training), the results of this analysis are equivalent to the results of a regression analysis where the dependent variable is a gain variable (Post-training scores minus Pre-training scores) and the predictor variables are Age, EF, and RV. This is the reason why both the question and the findings of this study can be conceptualized in terms of slopes (as in rate of change from pre- to post-training) and in terms of amount of change (as in gain from pre- to post-training), which is why we use these conceptualizations interchangeably throughout the paper.

[6] The analyses were conducted using the SAS 9.3 proc mixed procedure. The initial status (pre-training coefficients) from the mixed models are omitted because they are identical to the coefficients from the multiple regressions presented in Table 5, Panel 1.

Table 6
Mixed model analysis of the rate of change on the two outcome variables.

Predictors	Vitalism Rate of Change Coefficients	Fun Facts Rate of Change Coefficients
Intercept	0.36***	4.9***
Age	−0.03*	−0.02
Receptive Vocabulary	−0.01	0.57**
Executive Function	0.15**	0.002

Key: ~p ≤ .10.
* p ≤ .05.
** p ≤ .01.
*** p ≤ .001.

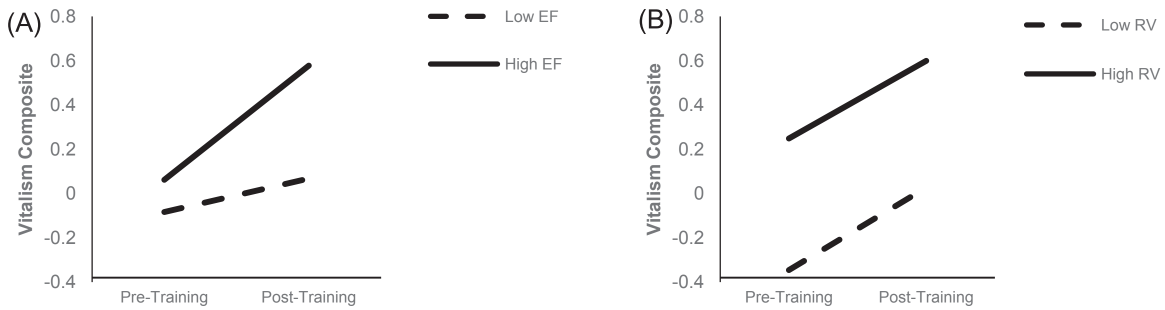

Fig. 1. Fitted growth trajectories for prototypical children on the composite **Vitalism** outcome variable (Fig. 1.1) and on the **Fun Facts** outcome variable (Fig. 1.2) (see below). Fig. 1.1 (A) Fitted growth trajectories on **Vitalism** for a prototypical child with an average age, average RV and low **EF** (10th percentile, dashed line) and high **EF** (90th percentile, solid line). (B) Fitted growth trajectories on **Vitalism** for a prototypical child with an average age, average EF and low **RV** (10th percentile, dashed line) and high **RV** (90th percentile, solid line).

every 1-unit difference in EF, on average, there is a 0.15 increase in the rate of change (amount of improvement) on the Vitalism scores.[7]

Because the correlations presented above showed that Fun Facts performance at post-training was correlated with Vitalism (see Table 3, Panel 2), it was important to check if children's performance on Fun Facts-Post, which represents children's ability to learn new facts, is a significant predictor of children's rate of change on Vitalism. Thus, in a separate mixed model, we included Vitalism as an outcome variable and Age, RV, EF, *and Fun Facts-Post* as predictor variables. The results were consistent with those presented in Table 6. Namely, neither RV nor Fun Facts-Post were significant predictors of the rate of change on Vitalism ($ps > .1$) and EF was ($p < .05$).

In summary, variance in the amount of improvement on Fun Facts is predicted by children's Receptive Vocabulary scores, but not by the composite Executive Function measure. Conversely, variance in the amount of improvement on the Vitalism composite score is not significantly related to Receptive Vocabulary scores (or to post-training Fun Facts scores), but it is predicted by the composite EF measure. These findings are a clean example of a double dissociation: EF is the only predictor of the rate of change on Vitalism scores, and RV is the only predictor of the rate of change on Fun Facts scores.

Fig. 1 is a visual representation of the results presented in Tables 5 and 6. The modeled pre-training and post-training points were computed by using the mixed model parameter estimates for children with low and high EFs, and for children with low and high RVs. Note that the points could also be computed by using the pre- and post-training regression coefficients presented in Table 5.

Had all children improved to the same extent as a result of the training (or had all children failed to improve), then there would have been no variance in the rate of change to be predicted by other predictors. However, as is evident in Table 6, not all children improved to the same extent on Vitalism. Fig. 1.1 displays modeled composite Vitalism scores of children with low (bottom 10th percentile) and high (top 90th percentile) EF scores (Fig. 1.1A) and low and high RV scores (Fig. 1.1B). We see that children with higher EF had steeper slopes between pre-training and post-training on the Vitalism scores, compared to children with lower EF (Fig. 1.1A). There was no such difference in the slopes between children with lower and higher RV scores (Fig. 1.1B). There was, however, a big effect of RV at pre-training that remained unchanged at post-training (i.e., children with higher RV scores scored much higher on the Vitalism composite than children with lower RV both at pre- and post-training), but, importantly, there was no significant interaction between RV and Training. Both children with low RV and children with high RV improved similarly as a result of the training. Taken together, these findings suggest that the variance in the *rate of change* on Vitalism is not associated with children's accumulated factual knowledge, nor it is associated with their ability to encode and retain new facts. Rather, they are

[7] In a separate analysis, we found that the two components of the composite EF variable individually predict the rate of change on Vitalism. That is, the Verbal Fluency composite is a significant predictor of the rate of change on Vitalism ($p < .05$) and Flanker is also a significant predictor ($p < .05$). When both predictors are entered in the model simultaneously, both are trending toward being significant ($ps \leq .1$).

consistent with the claim that the process that drives the *rate of change* on Vitalism (i.e., the construction of new theoretical knowledge) is associated with EFs.

An exception to this, however, might be the rate of change on Body Parts. Given that the training intervention included many facts about bodily function, it is likely that RV would be predictive of the rate of change on the Body Parts interview. To test this possibility, we repeated the same mixed model analysis presented above with Body Parts scores as an outcome variable and Age, RV, and EF as predictor variables of the rate of change. Only EF was trending toward being a significant predictor of the rate of change in Body Parts scores ($p = .06$) and RV was not significant ($p > .9$).

In summary, RV predicts variance in improvement in Fun Facts scores from pre- to post-training assessments, which indicates that RV reflects variability in the knowledge enrichment mechanisms that underlie learning fast mapped factual knowledge. In addition, RV predicts the Vitalism scores, both at pre-training and post-training (see Fig. 1). We have interpreted this latter finding in light of the obvious point that factual knowledge that is the outcome of knowledge enrichment processes alone is indubitably important to conceptual construction. Given these findings, and this interpretation, why did RV *not* predict degree of improvement on the Vitalism scores?

The fact that RV did not even predict degree of improvement on the Body Parts interview itself is a key to understanding this puzzle. The training intervention was not designed to teach to the test (to the Body Parts Interview), which has been in use in our labs for over 20 years. Rather, it was designed to support the creation of a framework vitalist theory. We did not consult the Body Parts interview in designing the training. Nonetheless, the training presented many new facts requiring knowledge enrichment alone that could have led to higher scores on the Body Parts interview. This is so because material covered when assessing children's understanding of the bodily function in supporting biological goals necessarily overlaps with the material designed to teach them that understanding. Thus, the training directly taught the answers to the questions: What are hearts, stomachs, brains, lungs and blood for? Why do we eat food? What happens to the food we eat? Why do we breathe air? What happens to the air we breathe? In spite of this, the overall Pre- to Post-training gains on the Body Parts Interview were modest (0.5 SD, compared to 4.9 SD for Fun Facts). In further exploring this issue, we asked whether the improvement on facts directly taught during training was more dramatic. It was not (0.6 SD rather than 0.5 SD; compared to 4.9 SD for Fun Facts). Furthermore, just as for the Body Parts Interview as a whole, an analysis of the predictors of the *improvement* on the directly taught Body Parts Facts found EF to be a significant predictor ($p = .02$), but not RV ($p = .38$), or age ($p = .59$).

Why were the Body Parts facts not fast-mapped? There are at least two plausible answers (not mutually exclusive). First, the training presented many facts, the long term encoding of which would likely interfere with each other. Second, the facts were not presented in isolation, but they were presented as part of a complex causal explanatory framework. The facts had to be understood to support the child's progress in mastery of that causal explanatory framework, but they did not have to be remembered. The training was designed to support *incorporating* already known facts and newly learned ones into a new causal structure. And it apparently did so. Not only did children improve on the Body Parts interview, but they improved on the Animism and Death interviews too. The learning of a new causal explanatory framework in particular is a very likely reason why the improvement on the Body Parts Interview, as well as that on all three interviews, was predicted by EF and not by RV.

The pattern of results for the modeled pre-training, rate of change, and post-training scores on the Fun Facts outcome variable is entirely different. Fig. 1.2 displays modeled Fun Facts scores of children with low (bottom 10th percentile) and high (top 90th percentile) EF scores (Fig. 1.2A) and low and high RV scores (Fig. 1.2B). First, EF has no effects. There is no effect of EF at pre-training assessment nor at post-training (Fig. 1.2A). The modeled pre-training Fun Facts score for a prototypical High EF child was similar to that of a prototypical Low EF child. Importantly, the large improvement between pre- and post-training was similar for a modeled child in the 10th percentile and one in the 90th percentile of our composite EF measure. Conversely, a modeled child in the 90th percentile of RV had steeper slopes between pre-training and post-training on Fun Facts, compared to a child in the 10th percentile (Fig. 1.2B). Indeed, a prototypical child with an average Age, average EF and high Receptive Vocabulary score (90th percentile) is predicted to have a score on Fun Facts at post-training that is 1.67 standard deviations higher than the score earned by a prototypical child with an average Age, average EF, and a low RV score (10th percentile). Thus, RV scores can predict who is poised to learn more from a training that includes a lot of new stand-alone facts, independent of other pieces of knowledge. In conclusion, these findings

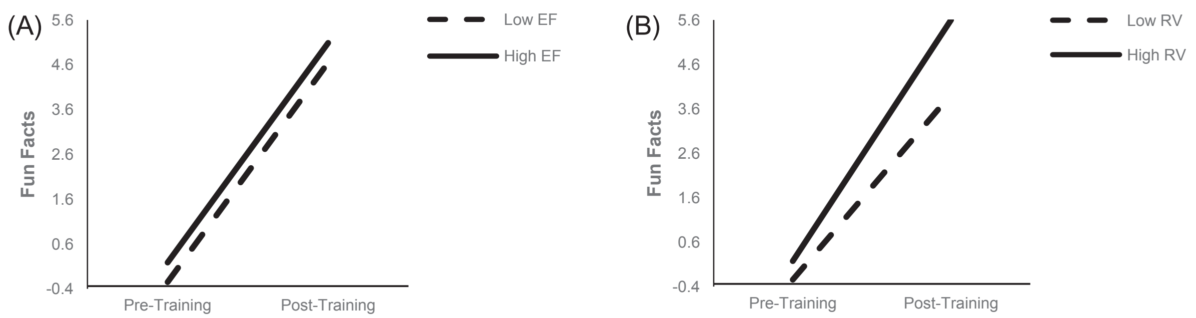

Fig. 2. (A) Fitted trajectories on **Fun Facts** for a prototypical child with an average age, average RV and low **EF** (10th percentile, dashed line) and high **EF** (90th percentile, solid line). (B) Fitted trajectories on **Fun Facts** for a prototypical child with an average age, average EF and low **RV** (10th percentile, dashed line) and high **RV** (90th percentile, solid line).

are consistent with the hypothesis that theory construction is supported by different domain-general cognitive resources than is learning new stand-alone facts. We turn to possible interpretations of this entire pattern of results in the general discussion, but note here that these results rule out one obvious interpretation of why children with high EF benefitted more from training than those with low EF. This is not because high EF children paid closer attention. If that were the case, they should have similarly done better in learning the Fun Facts.

A reviewer raised another possible interpretation of the fact that EF does not predict learning of Fun Facts. Might it be that these fast-mapped facts are fast mapped precisely because the child has no entrenched prior beliefs concerning them? Could it be that EFs predict improvement on the Vitalism scores only because learning vitalist theory requires inhibiting the enriched agency theory that is CS1? This account, which is very different from that advanced here, predicts that knowledge enrichment involving generic facts that conflict previously held beliefs might also draw on EFs. The present experiment was not designed to test this hypothesis, but two aspects of our data speak against it. First, the Fun Facts differed in terms of whether children had entrenched prior beliefs. For example, children had absolutely no idea about the size of a Blue Whale's tongue or about which bird lays the largest egg. On the other hand, other facts conflicted with entrenched knowledge derived from facts children took to be true of all animals (e.g., all ears are on heads, so a cricket's ears are on its head). Although at pre-training assessment children gave no answer to many Fun Fact questions, when they did so, the answers were predictable and consistent for some of the questions. For example, based on generalizations about animals, many children said that the ears of a cricket are on its head, that the heart of a shrimp is in its chest. For others there was no consistent response, and usually the response was simply "I don't know." We tested whether EFs would predict degree of change of the two types of facts differentially. We split the Fun Facts questions in two: (i) a set of questions that did elicit a dominant response among children (Prior Belief Facts, $n = 6$) and (ii) a set of questions that did not elicit consistent responses (Complete Ignorance Facts, $n = 4$). The criterion for being a prior belief fact was that at least 50% of the offered answers on a particular question were the same *incorrect* answer. We repeated the analyses reported above with the Prior Belief Facts and the Complete Ignorance Facts as separate outcome measures. The pattern of results was exactly the same for the two types of generic facts. For the Prior Belief Facts, RV was a significant predictor of the rate of change ($p = .001$) and EF was not ($p = .53$). For the Complete Ignorance Facts, RV was trending toward being a significant predictor of the rate of change ($p = .06$) and EF was not significant ($p = .93$).

A second finding also speaks against the hypothesis that EFs are implicated in learning only when inhibition of prior beliefs is involved. That is the finding that EFs but not RV predict the amount of improvement on Body Parts. Children's enriched core knowledge of intentional and causal agency is completely silent on the internal structure of the body and the functions of specific body parts. Indeed, preschoolers are ignorant about the function of body parts. Yet, EFs but not RV predicted the rate of change on Body Parts. As argued above, this is probably because the improvement seen on the Body Parts interview involves partial mastery of the framework theory of vitalism, and mastery of a new theoretical structure relies on learning mechanisms that employ EFs.

4. Discussion

The present study replicates and extends previous work in demonstrating that episodes of conceptual development involving fast-mapped knowledge enrichment alone differ from those involving conceptual construction. Most importantly, this study is the first to disassociate the domain-general cognitive resources that support these two types of learning. Two innovations made this possible. First, we directly compared tasks that diagnose progress toward a well-studied case of conceptual change in childhood (the Vitalism interviews) with tasks that diagnose mastery of generic facts the encoding of which does not require conceptual change (the Fun Facts). Second, we created training interventions targeting both and directly compared them with respect to the domain general capacities that were associated with progress of learning each.

As reviewed in the Introduction, previous research has shown that individuals with Williams Syndrome, a form of intellectual disability, never acquire a vitalist biology, despite their relatively intact ability to acquire new vocabulary and new facts (Johnson & Carey, 1998). Although this finding suggests that different domain-general resources are recruited for learning new facts than for acquiring new theories, Johnson and Carey's (1998) study did not specify which domain-general resources might be implicated in the two types of learning. The present results, along with those from Zaitchik et al. (2013, submitted for publication), begin to fill that gap.

To our knowledge, the present study is the first training study that explores the quantitative and qualitative distinctions between acquisition of new generic factual knowledge (the Fun Facts) and acquisition of a new framework theory (vitalist biology) requiring conceptual change for its construction. The distinction between these two aspects of knowledge acquisition is reflected in differences in the rate of acquisition, the interrelations among concepts within each, and the domain-general resources that support each.

4.1. Conceptual change and knowledge enrichment: distinct aspects of conceptual development

With respect to speed of acquisition, we found ten times greater advance in learning the Fun Facts than in learning the facts probed on the Body Parts interview, even though over the two training sessions the facts about body parts were the subject of a total of 40–60 min of instruction, repetition, and integration with respect to the tenets of vitalist biology, whereas the Fun Facts were presented just once during each of the training sessions. This suggests that the acquisition of new generic factual knowledge, formulated in terms of already available concepts, is indeed fast-mapped, and extends the phenomenon of fast mapping from lexical acquisition (Carey, 1978) and idiosyncratic fact learning (Markson & Bloom, 1997) to the learning of generic facts. Facts that require conceptual construction are not fast mapped.

Another way of differentiating conceptual change from knowledge enrichment is by examining the interrelations among concepts. Unlike episodes of knowledge enrichment in which the learner can acquire numerous unrelated facts, episodes of conceptual change always involve co-construction of interrelated concepts. The interrelatedness of CS2 concepts in the case of vitalism means that learning about one concept leads to learning about the other concepts, so long as the co-construction has already begun (i.e., a skeletal placeholder structure is in place) and some relevant knowledge formulated in terms of concepts common to CS1 and CS2 is in place. Previous research has shown that a training intervention with 5-year-olds concerning how bodily function supports life leads to progress on a death interview, even though death was not mentioned during training (Slaughter & Lyons, 2003). The present training intervention builds on Slaughter and Lyon's study, and both replicates and goes beyond their findings. Our training mentioned *neither* life nor death, yet, it resulted in far transfer on *both* the Animism and the Death interviews.

Further evidence concerning the interrelations among concepts derives from direct measures of coherence. Although the three vitalist interviews draw upon different aspects of vitalist theory, children's progress prior to training on any one interview predicted their progress on the other two, controlling for measures of factual knowledge (in this case both RV and Fun Facts; see also Slaughter & Lyons, 2003; Zaitchik et al., 2013, submitted for publication). The coherence among the interviews survives controlling for measures of accumulated factual knowledge. This is consistent with the hypothesis that conceptual co-construction draws on processes not involved in knowledge enrichment. Particularly strong (and novel) evidence for co-construction of the vitalist concepts *life, death,* and *the bodily machine*, derives from the finding that the coherence among the three Vitalism interviews was greater at post-training than at pre-training. During training, children learned *no new facts* about the animate/inanimate distinction (the focus of the Animism interview) or about what happens to the human body after death (the focus of the Death interview). Yet performance on the Body Parts interview was more related to performance on the Animism and Death interviews at post-training than at pre-training. This result provides strong support for two claims: first, vitalism is a coherent framework theory representing the inter-defined biological concepts of *life, death,* and *the bodily machine;* second, at least part of the skeleton of that framework theory is already in place in some six-year-olds at the outset of our training intervention.

The primary goal of this study was to explore and directly compare the domain general resources drawn on in the early stages of construction of a vitalist biology with those drawn on in the acquisition of factual knowledge. The results show that EFs (but not RV) predict the amount of improvement on Vitalism scores, controlling for all other variables, whereas RV (but not EFs) predicts the amount of improvement on Fun Facts scores, controlling for all other variables. This double dissociation provides strong evidence that different learning mechanisms underlie the two different kinds of conceptual development.

4.2. The role of factual knowledge acquisition in conceptual change

Children and adults learn much of what they know from the testimony of others, and this knowledge includes new facts stated in terms of concepts they already have. Acquiring such factual knowledge is very different from conceptual change, but such factual knowledge is critically important for conceptual change in multiple ways. Sometimes new facts motivate conceptual reorganization and change; sometimes they are placeholder structures that capture some of the CS2 structure; and sometimes they represent phenomena that are articulated in concepts that are common between CS1 and CS2, and thus are particularly suited to support the modeling processes that are important to the construction of CS2.

With respect to motivating conceptual reorganization, new facts can contradict an older accepted fact or prediction of CS1. Noticing such contradictions motivates attention, and sometimes motivates conceptual work. This is manifest in "why" questions (Callanan & Oakes, 1992) as well as in exchanges such as the following (from Carey, 1985b).

Child. That's funny, statues are not alive but you can still see them.
Mother. What's funny about that?
Child. Well, Grampa's dead and that's sad because we can never see him again.
Mother. Oh, I see. Well some things, like tables are chairs and statues are never alive so they can't die, but other things, like animals and plants and people, first they are alive and then they die, and that's sad because when they die they don't exist anymore, and we can never see them again.
Child (excitedly). Isn't that funny. Tables and chairs are not alive and we can still SEE them.

This exchange also illustrates that noticing a contradiction does not tell the child (or the scientist, see Carey, 2009; Wiser & Carey, 1983) how to resolve it. The 3-year-old's contradiction was articulated in terms of her undifferentiated concept *alive*. Importantly, she was in no position to understand her mother's attempt to point out to her that she had failed to differentiate the concept *dead* from the concept *inanimate.* Nonetheless, noticing the contradiction put into motion conceptual work to try to resolve it.

With respect to acquiring placeholder structures, the child can learn the fact that plants are alive, or that one can't ever see a dead person again, that dead people are buried under the ground, and their bodies decompose, without yet having the CS2 concepts *alive* and *dead*. These are part of placeholder structures, from the point of view of CS2, formulated in terms of relations between CS1 meanings for words like "alive" and "dead" at a point in development in which the child does not yet have the CS2 concepts. Nevertheless, they capture CS2 *interrelations* among concepts, and thus play a role in acquiring CS2 itself, as well as in constructing the central concepts of CS2.

The modeling processes that contribute to filling in the placeholder structures, which involve analogy, limiting case analyses, thought experimentation, and inductive inference, are constrained by knowledge of facts that are stated in terms of concepts held in *common* between CS1 and CS2. Examples include facts such as: "you need to eat to grow; the stomach is where the food you eat first goes," and many other facts. Indeed, most of the facts presented in the vitalism training in the present study fall in this category. Thus, although these facts are acquired through straightforward knowledge enrichment, they play crucial roles in the bootstrapping processes that underlie construction of the new CS2 concepts.

Several results from the present study support the importance of knowledge enrichment to conceptual change. Most strikingly, measures of factual knowledge are correlated with performance on the vitalism battery (see also Zaitchik et al., submitted for publication for converging results). Children who scored higher on RV also scored higher on Vitalism at pre-training and this advantage remained virtually unchanged at post-training (see Fig. 1.1B). There are two possible sources of variance on RV. One source is variance in the quality and quantity of input (from school, home environment, parental and teacher testimony) that imparts new factual knowledge. The second is variance in the capacity to encode and retain newly encountered factual knowledge. While not denying the important role of the first source of variance, the present experiment provides novel evidence that the latter source of variance is also at work here. Given the near floor effects on Fun Facts at pre-training, and given that all children were presented with the same quantity and quality of new facts during training, most of the variance on Fun-facts at post-training must be due to children's ability to *encode and retain new facts*. Importantly, the variance on the Fun Facts interview at *post-test* strongly predicted Vitalism scores, *both* at pre-training and at post-training, even after controlling for Age and RV. Of course, the correlation between Vitalism and Fun Facts is not because the Fun Facts induced any contradiction motivating revision of CS1, or new placeholder relation between core concepts of CS2, or new beliefs important to filling in CS2 placeholder structures. Rather, the Fun Facts scores capture children's variability in their efficiency at fast mapping and retaining new facts, their efficiency at knowledge enrichment. This ability, in turn, is important for knowledge restructuring and conceptual change. The main effect of RV on Vitalism scores most likely reflects a lifetime of accumulated factual knowledge that is the output of knowledge enrichment.

Although this is so, the *improvement* on the Vitalism battery is not due to fast mapped knowledge enrichment. The rate of change on Vitalism scores was not predicted by RV (or Fun Facts – see Section 3.8). As discussed above, the training effects on the Body Parts Interview, as well as the far transfer to the Animism and Death interviews, most likely reflects what children learned about a causal-explanatory framework according to which energy from the outside world is harnessed by a concerted work of the body organs, which work supports activity, growth, and health. Making sense of this information provided to them during the training sessions did not draw on their ability to encode and retain new facts, but rather on their capacity for set-shifting and inhibitory control.

4.3. The relationship between executive functions and conceptual construction

The present study found that variation in receptive vocabulary predicts variation in the rate of learning new facts; in contrast, variation in EF, specifically set-shifting and inhibition, the only EFs probed here, predicts variation in the rate of progression in the early stages of acquiring a vitalist theory of biology. There were no measures of individual differences in children's working memory, delay (or hot) inhibitory control, or other general cognitive abilities, such as fluid IQ. Thus, it is not clear if any of these domain-general abilities would have predicted improvement on the Vitalism or Fun Facts measures. However, as detailed in the introduction, Zaitchik et al. (submitted for publication) investigated the relationship between performance on the same Vitalism battery used here and a much larger battery of EF and IQ measures. These measures included inhibition and cued shifting, self-directed shifting, working memory, and fluid IQ (as assessed by a child version of progressive matrices). Zaitchik et al. found that Vitalism scores were *independently* predicted by inhibition and cued set shifting (Hearts and Flowers; Flanker) and by self-directed shifting (Verbal Fluency). In contrast, performance on the Vitalism battery was not related to working memory or to fluid IQ. Moreover, including working memory and fluid IQ in the regression models did not diminish the predictive power of inhibition and cued shifting or self-directed shifting. Zaitchik et al. did not investigate *improvement* on Vitalism scores as a function of training. Still, it stands to reason that if working memory and fluid IQ do not predict the overall measures of understanding of vitalist theory, which in turn reflects previous acquisition prior to testing, it is unlikely that they predict the rate of acquiring such understanding. Future work should investigate this question directly.

Why would executive functions, and specifically set-shifting and inhibition, be important for the acquisition of the vitalist theory of biology? Consider again the child whose initial theory of animals is an enriched theory of agency, according to which agency (i.e., intentional behavior, and self-propelled motion) is the essence of the concept *animal,* and is initially mapped to the word "alive." Sooner or later, that child will encounter testimony that plants, entities that do not exhibit intentional behavior and that do not move, are also alive. Some children might accept that testimony without even noticing that it conflicts with their other beliefs, whereas other children might notice the inherent conflict between believing that "alive" means "to move and do things" and believing that plants are alive. The child who notices the conflict may try to resolve it by appealing to the undifferentiated *alive/real* concept, saying that "plants are alive because you can see them." But that would bring to surface other conflicts, such as believing that rocks can be seen, but also that rocks are *not* alive (see dialog above concerning the conflict between the propositions "statues are not alive" and "we can see statues" that motivated a 3-year-old to seek clarification from her mother). Noticing conflicts among beliefs, attempting to resolve those conflicts by bringing other pieces of knowledge to bear, relating them to each other by using analogies and relational reasoning, generating new explanations while inhibiting others and flexibly shifting between them is a complex process that requires inhibitory control and set-shifting.

Children's conflicting beliefs, reflected in their responses on the Vitalism battery, were not explicitly pointed out in the vitalism training. Instead, the training emphasized the essential role of vitalist bodily functions in supporting growth, health, and activity, the last of which is a causally central property of animals within CS1. CS1 is an enriched agency theory, enriched by the undifferentiated concepts of life and death that the child has created to make sense of what will become biological phenomena. To assimilate the testimony from the training, children had to notice that activity, a causally central property of animals in CS1, is supported by vital energy coming from the outside world and that the function of various vital body organs is to harness that energy in support of bodily activity. They might have also recognized that a breakdown in the process would mean cessation of activity, and that the same vitalist processes of harnessing energy and material from the outside world that supports activity, also support growth, and that both animals and plants grow, but tables and cars don't. These realizations might lead some children to entertain the hypothesis that a breakdown in these physiological processes might mean cessation of life. Furthermore, extracting these causally related principles from the training requires both inhibition and set shifting—inhibition of established and prepotent modes of thinking (such as the agency construal of *life*), and shifting between different pieces of information and different frameworks of explanation.

Clearly, the construction and entrenchment of the framework vitalist theory of biology unfolds over a much longer time than the two training sessions provided here. It is aided, no doubt, by many episodes in which newly learned information is inconsistent with the child's current conceptualization of biological phenomena, episodes in which newly learned information adds to the coherent new set of concepts that are emerging. In any such episode, shifting and inhibition might be expected to play roles similar to those at play in the current study.

4.4. Implications concerning the nature of the learning mechanisms that underlie conceptual construction

That variablility in measures of set-shifting and inhibition predict variability in the slope of the pre-training/post-training scores is, as just described, consistent with the hypothesis that Quinian bootstrapping underlies the conceptual changes in biological concepts early in the elementary school years. The training intervention was designed to support Quinian bootstrapping, but the present study does not systematically investigate *why* the training worked, or even *that* the training intervention itself was important. This is because we had no control group who received the Vitalism battery twice, but no training, or different training regimes in between (but see Bascandziev and Carey, in preparation).

Nonetheless, it is unlikely that Quinian bootstrapping is the sole mechanism underlying conceptual construction. New computational primitives, not composable out of conceptual primitives currently available, may arise through many different mechanisms. For example, in Bayesian causal learning, new causal variables may be posited to make sense of the conditional probabilities observed among known variables. Sometimes, as in Quinian bootstrapping, initially perhaps nothing may be known about these variables—what they correspond to in the real world, and what the nature of the causal mechanism is that underlies the causal Bayes net (Tenenbaum, Griffiths, & Kemp, 2006). The process that fills those gaps may overlap considerably with Quinian bootstrapping—the content is captured initially entirely by conceptual role (in this case by a role in a Causal Bayes Net), and further processes are needed to provide additional sources of meaning. It is an empirical question whether EFs are implicated in the positing of new casual variables in the course of causal learning, and/or in the subsequent learning that supports conceptualizing these variables and conceptualizing the causal mechanisms involved.

In sum, the present work cannot be taken even as tentative evidence that EFs are required for *all* episodes of conceptual construction. Similar empirical questions would arise for any distinct mechanism that results in the construction of new conceptual primitives available to articulate thought.

4.5. The role of EFs in the expression of conceptual knowledge

We have interpreted our results as suggesting that EFs are drawn upon in the processes through which children *learn* the concepts of a vitalist biology. But there is a competing interpretation of these data, one that has some currency in the literature on the role of EFs in the developmental changes observed on explicit Theory of Mind tasks. It is possible that developing EFs contribute merely to the ability to *express* conceptual knowledge—conceptual knowledge that is already in place, either because of innate conceptual structures (as suggested in the Theory of Mind literature; see Baillargeon et al., 2010; Leslie, 1994; Leslie et al., 2004) or as the result of learning processes that did not, themselves, draw on EFs (as might be case for the concepts that articulate vitalist biology). By this hypothesis, variation in EFs predict variation in measures of conceptual understanding only because children with better EFs can better express their existing conceptual knowledge in the tasks that diagnose it.

There is no doubt that EFs are drawn upon in the expression of the vitalist theory. Many questions on the vitalism battery elicit prepotent responses that require inhibition. These are found in the Animism interview, with its prepotent agency responses, (see Goldberg & Thompson-Schill, 2009; Shtulman & Valcarcel, 2012; Tardiff et al., 2017). They may be found as well in the Death interview, where vitalist responses may well require the inhibition of religious responses (see Harris & Gimenéz, 2005, for evidence for two distinct framework theories of death, one deriving from vitalist biology and one derived from the conceptual system that frames spiritual and religious beliefs. It is thus possible that the relationship between EFs and performance on vitalism reflects *only* the role of EFs in expressing already existing conceptual knowledge (the expression alone hypothesis). On this hypothesis, all children may have learned equally well from the vitalism training in the present study, but only the children with higher EFs were able to express what they learned at post-training. Importantly, under this view, executive functions are not needed for theory construction.

While agreeing that executive functions are needed for the expression of already present conceptual knowledge, we believe that this expression alone hypothesis cannot fully explain the results of the present study. One important result that runs counter this claim is that the improvement on the Body Parts interview, which does not elicit any prepotent responses, was also marginally predicted by executive function, controlling for Age and RV, and there was no significant relationship between improvement on Body Parts and RV. Thus, although the questions about the function of body organs and what would happen if you didn't have them do not elicit prepotent responses, children with higher EFs improved more on this interview more than children with lower EF, likely because children with higher EFs were in a better position to assimilate the vitalism training and make the right inferences.

4.6. The relationship between fast-mapped receptive vocabulary and generic factual knowledge; implications for the learning mechanism underlying knowledge enrichment

It is well known that variance in measures of receptive vocabulary (RV) predict variance in measures of generic factual knowledge. For example, the WPPSI Receptive Vocabulary measure is highly correlated with the WPPSI Information test, a test of factual knowledge. A recent confirmation of this is Zaitchik et al.'s (submitted for publication) finding of a high correlation ($r = 0.66$, controlling for age) between scores on the K-BIT2 Verbal Knowledge Test (an RV measure) (Kaufman & Kaufman, 2004) and scores on the Woodcock-Johnson test of academic knowledge (Woodcock, McGrew, & Mather, 2001). The questions on the latter test vary widely in content difficulty, (e.g., from "What animal quacks?" to "In a period of inflation, what happens to the prices of goods and services available?"). Young children, including the 5- and 6-year-olds who participate in Vitalism studies, never make it to the latter item. Learning the facts probed on these measures require no conceptual construction. Indeed, the generic factual knowledge tapped on these instruments is fast-mapped. Young children already have concepts of animal kinds and the expectation that animals of different kinds make distinctive sounds. Similarly, older children have concepts of buying and selling, money, and prices for goods, and so can create the concept of inflation in terms of the same dollar amount buying less. There are at least three plausible, and not mutually exclusive, reasons why there is a close relationship between tests of accumulated information and tests of RV. One reason is that tests of information or generic factual knowledge *are*, partially, tests of vocabulary; one must know the meaning of words "quack" and "inflation" to answer these questions (not to mention "goods" and "services"). The second reason is that both vocabulary tests and tests of factual knowledge capture the shared variance in opportunity to learn new words or facts. Variation in input from parents (i.e., hearing more words/facts versus hearing fewer words/facts), quality of schooling, exposure to children's museums, zoos, science museums, enriching after school programs, and the like would lead to shared variance on both tests. Finally, the present results provide evidence for a third source of variance shared between vocabulary measures and measures of generic factual knowledge, and that is children's ability to create representations of new lexical items and new propositions, their ability to store them, and retrieve them from long-term memory. Although all children were exposed equally to the answers to the Fun Facts, those with higher RV improved more than did those with lower RVs between pre-training and post-training on Fun Facts. This reflects greater capacity of children with high RVs to form, store and retrieve (approximately 1 week later) the answers to the Fun Facts questions. This latter finding suggests that *the same learning mechanisms*, at least in part, underlie both learning new words and learning new generic propositions. A more direct test of this hypothesis, requires a comparison between the ability to learn new words (rather than the measure of accumulated receptive vocabulary that we used as a proxy for this ability in the present study) to the ability to learn new generic facts. Future research should investigate this relationship directly.

What might the relevant learning mechanism that underlies both lexical learning and the learning of new facts be like? How might it differ from the EF demanding mechanisms that underlie conceptual construction? Both conceptual construction and fast-mapped knowledge enrichment involve creating new representations, and both involve storing and retrieving them from long term memory. Speculatively, the differences are likely to lie in the process of creation of those representations: bootstrapping processes for conceptual construction, highly constrained domain-specific learning mechanisms for knowledge enrichment that results in fast mapping. In the case of both lexical learning and learning of facts, the constraints derive from knowledge of language. It seems likely that the variation common to RV and measures of information derives from variation in lexical, syntactic and semantic encoding abilities.

Any learning episode whereby new concepts or new propositions are composed from existing conceptual primitive is an episode of knowledge enrichment. Not all knowledge enrichment is fast-mapped. Hypothesis testing in which there is a huge, unconstrained, hypothesis space can be knowledge enrichment, but will not involve fast mapping. It is an open empirical question whether the domain general resources that distinguish fast-mapped knowledge enrichment from conceptual change also distinguish the latter from hard cases of knowledge enrichment, or whether EFs are drawn upon knowledge enrichment when it is hard. Also open is whether EFs are drawn upon generic factual knowledge learning that does not require conceptual construction for which there are strongly entrenched prior beliefs. The present Fun Fact Battery was not designed to test this hypothesis, but within the range of variation among the facts with respect to prior beliefs, there was no difference—prior belief facts and total ignorance facts patterned together: RV predicted learning and EF did not. Further research should more systematically address this issue.

There are other reasons that knowledge enrichment might be hard, and in each case one should explore what domain general conceptual resources are implicated. For example, knowledge enrichment involving large numbers of generic facts that interfere with each other, such as what the capitals of the 50 states are, is not easy. Also, procedural knowledge, like learning to spell, or learning to count, does not require conceptual construction, (although mastering how counting represents number, and perhaps acquiring the metalinguistic concepts of *word* or *sublexical sound unit* might), is not easy. Empirical work on the domain-general resources that underlie all of these cases of knowledge enrichment is necessary to explore how widely the pattern of results found here distinguishing conceptual change, on the one hand, from fast-mapped knowledge enrichment on the other, hold.

4.7. Concluding remarks

The acquisition of the theory of vitalism is hard. It takes children more than five years to fully develop a vitalist understanding of the biological world. One reason why this is so is that young children have an entrenched framework theory through which they understand animals—the agency theory—and this does a very good job making sense of much of what children care about with respect to animals. A second reason is that theory construction is hard. The bootstrapping processes that are implicated in this episode of conceptual change go beyond knowledge enrichment alone (Johnson & Carey, 1998; Zaitchik et al., 2013, submitted for publication). The findings of the present study provide a new and different kind of evidence for the difference between conceptual change and fast-mapped knowledge enrichment. Importantly, they show that EF is important for conceptual change but not for knowledge enrichment that results in fast mapping. These findings suggest that at least one reason why EF predicts school readiness and school success is that it is implicated in conceptual change. Children's acquisition of the conceptual systems that were culturally constructed over centuries, and that are the target of formal education in all subject matters, but most notably in STEM subjects within the mathematics and science curriculum, requires conceptual change.

Acknowledgments

This work was funded with support from the National Science Foundation #1247396, Co-PIS: Susan Carey and Deborah Zaitchik. We are grateful to the children and their families who participated in this study and we thank the administrators and teachers at the Jewish Community Day School in Watertown, MA and the Westborough Public Schools in Westborough, MA.

Appendix A

A.1. The four interviews with detailed scoring criteria

A.1.1. Animism interview

1. **What does it mean to be alive, to be a living thing?**
   3 points: Biological processes *(growth, death, life cycle)*
   2 points: Biologically relevant behavior *(breathing, eating)*
   2 points: Means you will die
   1 point: Movement, activity, or 'not dead', health
   0 points: Don't know, lack of response, is an animal/human, existence, emotions, meaning of life
2. **Can you name some things that are alive, that are living things?**
   2 points: People, plants, animals, either in general or specific examples
   1 point: People and/or animals but not plants
   0 points: Don't know, lack of response
   −1 point subtracted for inanimate objects
3. **Can you name some things that are NOT alive, that are NOT living things?**
   2 points: Any inanimate object
   1 point: Dead people, dead animals, or dead plants; dinosaurs or extinct animals
   0 points: Don't know, lack of response
   −1 point subtracted for imaginary entities
   −1 point subtracted for animals or plants
4. **JUDGMENTS: For each object, ask: "Is [an] x alive, is it a living thing?"**

object	Yes	No
a Mountain		
b Bell		
c the Sun		
d Wind		
e Table		
f Fly		
g Car		
h Fire		
i Cat		
j Pencil		
k Bird		
l Tree		
m Snake		
n Bicycle		
o Watch		
p Flower		
q Airplane		
r Lamp		
s Cloud		
t Rain		

"An x is alive, but it's not a living thing" is a wrong answer!
Overall:
4 points: Yes to all animals and both plants; no to natural phenomena and artifacts
3 points: Yes to all animals, no to one/both plants, natural phenomena, and artifacts
2 points: Yes to all animals and one or more natural phenomena; no to artifacts
1 point: Yes to all animals and one or more artifacts

5. **Probe a single judgment for each category: animal, plant, natural kind, artifact, as follows: "You said (an) x is/is not alive, is/is not a living thing. Can you tell me how you know?**
   a animal (cat)
   b plant (tree)
   c natural kind (Sun)
   d artifact (lamp)

2 points: For animal, a biological response (bodily responses, requirements, or relevant internal body parts) e.g. "a cat is alive because it breathes". For artifacts, man-made origins or inorganic composition, e.g. "an airplane is not alive because people made it."
1 point: Autonomous motion, e.g. "a cat is alive because it does things by itself," "an airplane is not alive because it can't move by itself"
0 points: Incorrect responses, movement, activity, is an animal, external body parts, e.g. "a cat is alive because it moves and has feet"

−1 point: Existence, e.g. "a cat is alive because I see it"

For each of the following questions, points were attributed in line with the above schema so long as the specific justifications were not previously mentioned, e.g. a question 3 response "a table is not alive because it doesn't breathe" would get no points if it followed a question 1 response that "a cat is alive because it breathes", but would get 2 points if it followed a question 1 response that "a cat is alive because it grows."
The points are summed up to create a composite Animism score.

A.2. Death interview

1. **What does it mean to die?**
2. **Can you tell me some things that die?**
3. **What happens to a person's body when they die?**

With respect to the following properties, responses to questions 1 and 3 considered as a single response: possible score of 3.
1 point for each of the following: (a) Breakdown of bodily function *(*e.g. *"stop breathing")* (b) decomposition *(*e.g. *body rots)*, or (c) Death as end of life *(*e.g. *not alive anymore")*
0 points: Behavioral responses, burial, religious interpretations of death, or death as opposite of life *(*e.g. *"not alive")*

−1 point: religious responses only, e.g. "go to heaven" to Q3.

Points are summed as follows for question #2:
2 points: All living things must eventually die
1 point: People and/or animals, either general or specific examples
1 point: Plants, either general or specific examples
0 points: Someone or something that is already dead *(*e.g. *George Washington")*, dinosaurs, or extinct animals
−1 point: Imaginary entities
−1 point: Inanimate objects

4. **When a person is dead, does he need to eat food?**
5. **When a person is dead, does he need to pee?**
6. **When a person is dead, does he need to sleep?**

0 points: No
−1 point for each Yes or don't know response, or for sleep *"no because they're already asleep"*

7. **When a person is dead, does he feel bad that he died?**
8. **When a person is dead, does he miss his friends?**
9. **When a person is dead, does he think about things?**

0 points: No
−1 point for each Yes or don't know response, unless it is mentioned in reference to the soul and not the body

10. **What might cause a person to die? [For one cause listed, probe:] Why would [x] cause a person to die?**

3 points: Cessation of biological function *(heart stop beating)*
2 points: Illness or old age
1 point: Specific avoidable cause *(getting shot)*
The points are summed up to create a composite Death score.

A.3. Body parts interview

1a. What is your brain for?
1b. What would happen if somebody didn't have a brain?
2a. What is your heart for?
2b. What would happen if somebody didn't have a heart?
3a. What are your lungs for?
3b. What would happen if somebody didn't have lungs?
4a. What are your eyes for?
4b. What would happen if somebody didn't have eyes?
5a. What is your stomach for?
5b. What would happen if somebody didn't have a stomach?
6a. What is your blood for?
6b. What would happen if somebody didn't have blood?

2 points: relating organ's function to another bodily function or explaining how the specific function they identify supports a biological goal *("the heart pumps the blood around to different parts of the body")* – *"the brain is the boss of your body"* gets 2 points
1 point: One organ – one function *("brain is for thinking")*
0 points: Don't know, lack of response, non-biological function *("lungs are for talking")*
For each organ, 1 additional point was given for mentioning the organ is needed to stay alive *("if we didn't have a heart, we would die")*.
A biological response that is correct, but not for the specific organ, e.g. *"heart is for breathing air and oxygen that we need to stay alive"* gets 2 points!

7a. Why do we eat food?

2 points: Specific biological response (energy, nutrients, growth, nourishment) – "so we don't starve to death" gets 2 points.
1 point: Non-specific biological response *(strength, health, life)* – *"so we don't starve"* gets 1 point
0 points: psychological response *("because we are hungry"),* don't know, lack of response.

7b. What happens to the food we eat?

3 points: Relating food to the functioning of the body ("the heart passes the food to parts of the body that need it")
2 points: Biological response ("it becomes poop", "it gets digested")
1 point: Purely mechanical response ("it goes into the stomach", "it gets mushed up")
0 points: Don't know, lack of response

7c. Do you need a brain to be able to eat?

1 point: Yes
0 points: No, don't know

8a. Why do we breathe air?

2 points: Biological response *("to bring oxygen")*
1 point: Behavioral response *("to talk")* or response not specific to organ *("to live")*
0 points: Don't know, lack of response

8b. What happens to the air we breathe?

2 points: Biological response *("air changes into energy")*

1 point: Purely mechanical response *("it goes in and out")*
0 points: Don't know, lack of response

8c. Do you need a brain to be able to breathe?

1 point: Yes
0 points: No, don't know
The points are summed up to create a composite Body Parts score.

A.4. Fun facts interview

1. **Do you know where the ears of a cricket are? (1 point)**
   Answer: They are on their legs, just below the knee.
2. **Do you know how dolphins sleep? (1 point)**
   Answer: They sleep half-awake, with half of their brain sleeping, while the other half is awake and one eye is closed and the other eye is open.
3. **Do you know which bird lays the biggest eggs? (1 point)**
   Answer: Ostrich.
4. **Do you know how many hearts does an octopus have? (1 point)**
   Answer: Three.
5. **Do you know how big the tongue of a blue whale is? (1 point)**
   Answer: It is as big as a fully grown African elephant.
6. **Do you know what color is an octopus's blood? (1 point)**
   Answer: Blue.
7. **Do you know where the heart of a shrimp is? (1 point)**
   Answer: The back of its head.
8. **Do you know what's the fastest land animal in the world? (1 point)**
   Answer: Cheetah
9. **Do you know how many legs does a spider have? (1 point)**
   Answer: Eight
10. **How many stomachs does a cow have? (1 point)**
    Answer: Four.

The points are summed up to create a composite Fun Facts score.

Appendix B

B.1. Vitalism Training Script

The training began with a question: "Do you know what's inside the body? Tell me some things that are inside our bodies. Can someone show me where the heart is? Can someone show me where the lungs are? Can someone show me where the stomach is? And how about the intestines? And do you know where the brain is? Here is a tricky one. Do you know where the blood is? That's right… the blood is everywhere… it's in all parts of the body."

Next, the experimenter introduced the 3-D anatomical model and said: "See, our bodies have different parts and this is how they look like. What do you think, what are these (take the lungs out)? That's right; this is how the lungs look like." After children inspect how the lungs look like, the experimenter asked: "Who wants to keep/be responsible for the lungs? Okay, you keep them now and make sure to pay special attention when we talk about the lungs. Do you know what this is (the heart)? That's right; that's how the heart looks like. Who wants to be responsible for the heart? Do you know what this is (the stomach)? That's right; that's how the stomach looks like. And these – do you know what these are (the intestines)? Do you know what these are (the kidneys)? These are called kidneys. And where is the brain? That's right. It's inside the head. Let's open the head… here is how the brain looks like." All organs were assigned to children and all children were responsible for at least one organ. "I want you to remember as much as you can about the organ you have. I will tell you how each organ works and why they are important and I want you to listen carefully."

Next, the experimenter said: "First, let me tell you something about some important things, such as food and air. Food and air are very important, because they give us energy to move, because they help us grow and they help us stay healthy. What do cars need to move? That's right; cars need gas. We need food and air to move. And we can move all parts. We can move our feet (the experimenter demonstrated each movement), we can move our hands, we can move our heads, we can move our shoulders and that is because the energy from the food and air gets to all parts of our bodies. Can you try and move your hands – see you need energy to move your hands. Can you move your feet? See you need energy to move your feet and the energy from food and air gets to all parts of your body. But there is another thing why food is important. Do cars grow up? No, but we do and food is important for growing up. You know when you put blocks on top of each other and make a tower that grows bigger and bigger. The body grows when you eat food. The food is like building blocks. It helps you grow bigger and stronger. And all parts of our bodies grow – your feet grow and your legs grow and your arms grow… and that's because the food gets to all parts of your body. Also, the food that you eat has ingredients that keep you healthy. If some part of the body does not get food or oxygen, it will get very, very sick. That's why it is important the food and the air to get to all parts of your body.

So, how do food and oxygen get to all parts of the body? That's what we'll learn about today. Can someone tell me what happens to the food you eat when it's in your mouth? That's right. When we eat food, we first chew it with our teeth and that cuts it into very little pieces. What happens next? Where does the food go when you swallow? That's right, it goes into the stomach. Who is responsible for the stomach? Good. So tell me everyone, what happens to the food when it's in the stomach? It gets smushed and mixed and it gets broken down *into even smaller pieces*. And where does it go from there? It goes into the intestines. Who is responsible for the intestines? Good. And what happens to the food when it is in the intestines? When the food is in the intestines it gets broken down *into even smaller pieces. Soooo small* that you can't even see them. And where does the food go from there? Does anyone know? It goes into the blood. Where can you find blood inside the body? You can find it everywhere. The blood goes to *all* parts of the body. And you know how it gets there? It travels around your body in tubes called blood vessels. Can you see some blood vessels in this plastic model? Remember, they look like tubes! That's right. Those are some blood vessels and as you can see, some are *very* big, some are big, some are small, and some are very, very small.

Do you know, what moves the blood around these blood vessels? The heart moves the blood. Who is responsible for the heart? Okay, so listen carefully. The heart is like a pump. It pumps, and pumps, and pumps (show with a gesture: squeeze the hand). Can you feel your heart beating? That beating is your heart pumping blood… and what does that do? That's right. It moves your blood around *all* parts of your body. And do you remember what does the blood carry? It carries the tiny pieces of food and it takes them around all parts of your body. It's like a bus that takes passengers to different parts of the city. So the blood takes the food around the different parts of your body, just like buses take people to different parts of the city. And why is that important. Do you remember? It's important because the food brings energy that you need to move your feet, your legs, your hands, your head, and it helps all parts of the body to grow, and it helps you to stay healthy.

You know what else. The blood needs to be clean and the kidneys help the blood stay clean. Who is responsible for the kidneys? Good. So, listen carefully. Do you remember where the kidneys in our body are? Very good. The kidneys are like a filter. The blood moves through the kidneys and it comes out clean. It's like a carwash. The dirty cars go in and they come out clean. The dirty blood goes through the kidneys and comes out clean. And do you know what happens to the air we breathe? It goes in our lungs. Everyone, take a deep breath. Can you feel how your lungs get filled with air? Do you know what happens in your lungs? *The air you breathe in gets inside your blood.* And then what happens? That's right. The blood takes the air to all parts of your body and that gives you *energy* to move your body and it keeps you healthy. Do you know what tells the heart how fast to beat? And the lungs how fast to breathe? Your brain does. Who is responsible for the brain? Good. So listen carefully everybody. The brain is the boss of your body. Nothing would work without the brain. You need the brain to think and see and to feel happy or sad, but you also need your brain to move your arms and legs, and you need the brain for your heart to beat and the blood to move around. And to be healthy and to work fine, the brain needs food and oxygen, just like any other part of your body.

Next, the experimenter gave the answers to the fun facts questions while showing them pictures and drawings that illustrated the answers. Finally, at the end of the session, children were asked to a) explain what organ they have, b) show where that organ goes (on the anatomical model), and c) say what the function of that organ is. For example, the experimenter called on children and said: "what do you have there? Great! That's right. And can you show us where does the ____ go here on the model? Great! That is correct. Can you tell us what the ____ is for and why it is important?" If the child did not remember, then the experimenter asked the group to help answer the question. Once the question was answered, the experimenter repeated the relevant information about that organ.

References

Ahn, W., & Kim, N. S. (2000). The role of causal features in categorization: An overview. In D. L. Medin (Ed.). *Psychology of learning and motivation* (pp. 23–65). New York: Academic Press.

Astuti, R., & Harris, P. L. (2008). Understanding mortality and the life of the ancestors in rural Madagascar. *Cognitive Science, 32*, 713–740.

Astuti, R., Solomon, G. E., & Carey, S. (2004). Constraints on conceptual development: A case study of the acquisition of folkbiological and folksociological knowledge in Madagascar. *Monographs of the Society for Research in Child Development, 69*, 1–161.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences, 14*, 110–118.

Baker, S. T., Gjersoe, N. L., Sibielska-Woch, K., Leslie, A. M., & Hood, B. M. (2011). Inhibitory control interacts with core knowledge in toddlers' manual search for an occluded object. *Developmental Science,* 270–279.

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind.* Oxford, UK: Oxford University Press.

Bascandziev, I., Powell, L. J., Harris, P. L., & Carey, S. (2016). A role for executive functions in explanatory understanding of the physical world. *Cognitive Development, 39*, 71–85.

Benson, J. E., Sabbagh, M. A., Carlson, S. M., & Zelazo, P. D. (2013). Individual differences in executive functioning predict preschoolers' improvement from theory-of-mind training. *Developmental Psychology, 49*, 1615–1627.

Block, N. J. (1986). Advertisement for a semantics for psychology. In P. A. French (Ed.). *Midwest studies in philosophy* (pp. 615–678). Minneapolis: University of Minnesota.

Callanan, M. A., & Oakes, L. M. (1992). Preschooler's questions and parents' explanations: Causal thinking in everyday activity. *Cognitive Development, 7*, 213–233.

Carey, S. (1985a). Are children fundamentally different thinkers and learners from adults? In S. F. Chipman, J. W. Sega, & R. Glaser (Eds.), *Thinking and learning skills: Vol. 2*Hillsdale, NJ: Erlbaum (Reprinted by Open University Readings in Cognitive Development (1985)).

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

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

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference, 15*, 17–29.

Carey, S. (1978). The child as word learner. In J. Bresnan, G. Miller, & M. Halle (Eds.). *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.

Carey, S. (1999). Sources of conceptual change. In E. K. Cholnick, K. Nelson, S. A. Gelman, & P. Miller (Eds.). *Conceptual development: Piaget's legacy* (pp. 293–326). Hillsdale, NJ: Erlbaum.

Carey, S. (2015). Why theories of concepts should not ignore the problem of acquisition. In E. Margolis, & S. Laurence (Eds.). *Concepts: New directions* (pp. 415–454). Cambridge MA: MIT Press.

Carey, S., Zaitchik, D., & Bascandziev, I. (2015). Theories of development: In dialog with Jean Piaget. *Developmental Review, 38*, 36–54.
Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032–1053.
Chatham, C. H., Frank, M. J., & Munakata, Y. (2009). Pupillometric and behavioral markers of a developmental shift in the temporal dynamics of cognitive control. *PNAS, 106*, 5529–5533.
Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics, 50*, 66–71.
Contento, I. (1981). Children's thinking about food and eating: A Piagetian-based study. *Journal of Nutrition Education, 13*, 86–90.
Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task-switching. *Neuropshychologia, 44*, 2037–2078.
Devine, R. T., & Highes, C. (2014). Relations between false belief understanding and executive function in early childhood: a meta-analysis. *Child Development, 85*, 1777–1794.
Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*, 1387–1388.
Diedehofen, B., & Musch, J. (2016). Cocron: A web interface and R package for statistical comparison of Cronbach's alpha coefficients. *International Journal of Internet Science, 11*, 51–60.
Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement, 11*, 93–103.
Ford, M. E., & Keating, D. P. (1981). Developmental and individual differences in long-term memory retrieval: Process and organization. *Child Development, 52*, 234–241.
Gallistel, R., Brown, A., Carey, S., Gelman, R., & Keil, F. (1991). Lessons from animal learningfor the study of cognitive development. In S. Carey, & R. Gelman (Eds.). *The epigenesis of mind: Essays in biology and cognition* (pp. 3–36). Hillsdale, NJ: Erlbaum.
Gelman, S. (2005). *The essential child: Origins of essentialism in everyday thought.* Oxford: Oxford University Press.
Gelman, S., Coley, J., & Gottfried, G. (1994). Essentialist beliefs in children: The acquisition of concepts and theories. In L. Hirschfeld, & S. Gelman (Eds.). *Mapping the mind: Domain specificity in cognition and culture*. New York: Cambridge University Press.
Gentner, D. (2002). Analogy in scientific discovery: The case of Jonannes Kepler. In L. Magnani, & N. J. Nersessian (Eds.). *Model-based reasoning: Science, technology, values* (pp. 21–39). New York: Kluwer Academic/Plenum.
Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences, 7*, 287–292.
Goldberg, R. F., & Thompson-Schill, S. L. (2009). Developmental 'roots' in mature biological knowledge. *Psychological Science, 20*, 480–487.
Harris, P. L., & Gimenéz, M. (2005). Children's acceptance of conflicting testimony: The case of death. *Journal of Cognition and Culture, 5*, 143–164.
Hatano, G., & Inagaki, K. (1994). Young children's naïve theory of biology. *Cognition, 50*, 171–188.
Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61–83.
Inagaki, K., & Hatano, G. (1993). Young children's understanding of the mind-body distinction. *Child Development, 64*, 1534–1549.
Inagaki, K., & Hatano, G. (2002). *Young children's naïve thinking about the biological world.* New York: Psychology Press.
Johnson, S. C., & Carey, S. (1998). Knowledge enrichment and conceptual change in folk biology: Evidence from people with Williams syndrome. *Cognitive Psychology, 37*, 156–200.
Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman assessment battery for children* (second edition). Circle Pines, MN: American Guidance Service.
Laurendau, M., & Pinard, A. (1962). *Causal thinking in the child: A genetic and experimental approach.* New York: International Universities Press.
Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences, 8*, 528–533.
Leslie, A. M. (1994). ToMM, ToBy, and Agency: Core architecture and domain specificity. In L. Hirschfeld, & S. Gelman (Eds.). *Mapping the mind: Domain specificity in cognition and culture* (pp. 119–148). Cambridge, UK: Cambridge University Press.
Leslie, A. M., & Polizzi, P. (1998). Inhibitory processing in the false-belief task: Two conjectures. *Developmental Science, 1*, 247–253.
Markson, L., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature, 385*, 813–815.
McCloskey, M., Caramazza, A., & Green, B. (1980). Curvilinear motion in the absence of external forces: Naïve beliefs about the motion of objects. *Science, 210*, 1139–1141.
McCloskey, M. (1983). Naive theories of motion. In D. Gentner, & A. Stevens (Eds.). *Mental models* (pp. 229–324). Hillsdale, NJ: Erlbaum.
Moses, L. J., & Tahiroglu, D. (2010). Clarifying the relation between executive function and children's theory of mind. In B Sokol, U. Muller, J. Carpendale, A. Young, & G. Iarocci (Eds.). *Self- and social-regulation: Exploring the relations between social interaction, social understanding, and the development of executive functions*. New York: Oxford University Press.
Nersessian, N. J. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In R. N. Giere (Ed.). *Cognitive models of science* (pp. 3–45). Minneapolis, MN: University of Minnesota Press.
Osorio, A., Raquel, C., Sampaio, A., Garayzabal, E., Martinez-Regueiro, R., Goncalves, O., ... Fernandez-Prieto, M. (2012). How executive functions are related to intelligence in Williams syndrome. *Research in Developmental Disabilities, 33*, 1169–1175.
Perner, J. (1991). *Understanding the representational mind.* Cambridge, MA: MIT Press.
Piaget, J. (1929). *The child's conception of the world.* London: Routledge and Kegan Paul.
Porter, M. A., Coltheart, M., & Langdon, R. (2007). The neuropsychological basis of hypersociability in Williams and Down syndrome. *Neuropsychologia, 45*, 2839–2849.
Powell, L. J., & Carey, S. (2017). Executive function depletion in children and its impact on theory of mind. *Cognition, 164*, 150–162.
Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black, & W. F. Prokasy (Eds.). *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.
Rhodes, S. M., Riby, D. M., Park, J., Fraser, E., & Campbell, L. E. (2010). Executive neuropsychological functioning in individuals with Williams syndrome. *Neuropsychologia, 48*, 1216–1226.
Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind. *Psychological Science, 17*, 74–81.
Saxe, R., & Carey, S. (2006). The perception of causation in infancy. *Acta Psychologica, 123*(1–2), 144–165.
Senju, A., Southgate, V., Snape, C., leonard, M., & Csibra, G. (2011). Do 18-month-olds really attribute mental states to others? A critical test. *Psychological Science, 22*, 878–880.
Setoh, P., Wu, D., Baillargeon, R., & Gelman, R. (2013). Young infants have biological expectations about animals. *Proceedings of the National Academy of Sciences, USA, 110*(40), 15937–15942.
Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition, 124*, 209–215.
Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modelling change and event occurrence.* New York: Oxford University Press.
Slaughter, V., Jaakkola, K., & Carey, S. (1999). Constructing a coherent theory: Children's biological understanding of life and death. In M. Siegel, & C. C. Peterson (Eds.). *Children's understanding of biology and health* (pp. 71–98). Cambridge, UK: Cambridge University Press.
Slaughter, V., & Lyons, M. (2003). Learning about life and death in early childhood. *Cognitive Psychology, 46*, 1–30.
Strevens, M. (2000). The essentialist aspect of naïve theories. *Cognition, 74*, 149–175.
Tardiff, N., Bascandziev, I., Sandor, K., Carey, S., & Zaitchik, D. (2017). Some consequences of normal aging for generating conceptual explanations: A case study of vitalist biology. *Cognitive Psychology, 95*, 145–163.
Taverna, A. S., Waxman, S. R., Medin, D. L., Moscoloni, N., & Peralta, O. A. (2014). Naming the living things: Linguistic, experiential and cultural factors in Wichí and Spanish speaking children. *Journal of Culture and Cognition, 14*, 213–233.
Taverna, A., Waxman, S., Medin, D., & Peralta, O. (2012). Core-folkbiological concepts: New evidence from Wichí children and adults. *Journal of Cognition and Culture, 12*, 339–358.
Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences, 10*, 309–317.
Waxman, S., Medin, D. L., & Ross, N. (2007). Folk-biological reasoning from a cross-cultural developmental perspective: Early essentialist notions are shaped by cultural beliefs. *Developmental Psychology, 43*, 294–308.
Wellman, H. M. (1990). *The child's theory of mind.* Cambridge, MA: MIT Press.
Wellman, H. M., & Gelman, S. (1992). Cognitive development: Foundational theories of core domains. *Annual Review of Psychology, 43*, 337–375.
Wiser, M., & Carey, S. (1983). When heat and temperature were one. In D. Gentner, & A. Stevens (Eds.). *Mental models* (pp. 267–297). Hillsdale, NJ: Erlbaum.
Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1–34.
Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson® III Test.* Itasca, IL: Riverside Publishing Company.

Zaitchik, D., Tardiff, N., Bascandziev, I., & Carey, S. (2017). The role of executive functions and fluid intelligence in young children's reasoning about biology (submitted for publication).
Zaitchik, D., Iqbal, Y., & Carey, S. (2013). The effect of executive function on biological reasoning in young children: An individual differences study. *Child Development, 85*, 160–175.

Further reading

Rhodes, S. M., Riby, D. M., Fraser, E., & Campbell, L. E. (2011). The extent of working memory deficits associated with Williams syndrome: Exploration of verbal and spatial domains and executively controlled processes. *Brain and Cognition, 77*, 208–214.