Children’s developing representations of mental life: Changes in conceptual structure between 4-9 years of age

Draft updated 2018-09-17

Kara Weisman, Carol S. Dweck, & Ellen M. Markman

# Abstract

XX

# Introduction

XX

In this paper, we first introduce a child-friendly version of Weisman et al.’s (2017) experimental paradigm, validate it among US adults, and then use this paradigm to conduct an initial exploration of this conceptual structure and mental capacity attributions among US children, focusing on 7- to 9-year-old children (Study 1). Next, we replicate our findings with 7- to 9-year-old children using a briefer experimental paradigm, and use this briefer paradigm to assess the earlier development of this conceptual structure, focusing on 4- to 6-year-old children (Study 2). We then develop an even simpler version of the paradigm, using more basic vocabulary and a streamlined experimental protocol, with the aim of validating and refining our understanding of conceptual structure among 4- to 6-year-old children, as well as exploring how younger children’s attributions of different aspects of mental life might vary with age (Study 3). Finally, to supplement the group-level analyses picture of development provided by the exploratory factor analyses for Studies 1-3, we re-analyze these data using a novel, participant-level approach that allows us to observe how these conceptual structures might evolve and unfold continuously over development (“Continuous development at the participant level: A re-analysis of Studies 1-3”).

# Overview of Methods and Analyses

Studies 1-3 used very similar experimental paradigms and were designed with the same analyses in mind. Here we provide an overview of the methods and planned analyses for all of these studies; any details that varied across studies are supplied in the “Methods” section of each of the studies.

## General methods

Following Weisman et al. (2017), these studies were designed with the goal of focusing participants’ attention on the similarities, differences, and relationships among various mental capacities.

To this end, each participant was either randomly or pseudo-randomly assigned to assess 1-2 target entities (e.g., a beetle, a robot, a goat, etc.) on a wide range of sensory, perceptual, emotional, social, cognitive, and other mental capacities, ranging in number from 18-40 across studies and presented in either a random or a pseudo-random (counterbalanced) order. Participants were presented with a vivid, full-color photograph of their assigned target in a naturalistic context (e.g., a beetle on a leaf; a robot in an office; a goat in a grassy field), which they had access to throughout the study.

On each trial, participants were asked a question of the form Do you think a [target] can [do X]? (e.g., “Do you think a beetle can feel happy?”; Studies 1-2) or Can [targets] [do X]? (Study 3) (e.g., “Can beetles feel happy?”). Participants responded on a three-point scale (no, coded as 0; kinda, coded as 0.5; or yes, coded as 1). Although a three-point scale is not optimal for factor analyses, pilot testing suggested that it was critical in allowing children to move fast enough through the study to answer all questions, and maintaining this within-subjects design was our top priority for the planned analyses.

See the Methods section for each study for details of the particular target entities and mental capacities included in each study, as well as the materials and physical setup.

## Primary analysis: Exploratory factor analysis (EFA)

Our primary goal in conducting these studies was to uncover a set of latent constructs that might plausibly have given rise to the observed correlations among mental capacity attributions in each group of participants. As such, our primary planned analysis for all studies was an exploratory factor analysis (EFA). Following Weisman et al. (2017), we interpreted each of the constructs (“factors”) as corresponding to a fundamental component of mental life, according to this group of participants; by extension, we consider the full set of factors for each sample to represent to the overall “conceptual structure” underlying mental capacity attributions for this group of participants.

For all EFAs, we used ordinary least squares to find the minimum residual solution, using the “psych” package for R (Revelle, 2018). Here we focus on results using Pearson correlations using pairwise complete observations. (See Online Supplementary Materials [OSM] for solutions using polychoric correlations, which are better suited to handle responses on a three-point scale but, to our eyes, tended to over-fit our data by suggesting that we should retain many factors that each accounted for only a small amount of the shared variance.)

In order to determine how many factors to retain, we examined the results of three factor retention protocols: (1) Parallel Analysis, which compares the observed correlation structure to the correlation structure arising from random datasets of the same size; (2) Minimizing the Bayesian Information Criterion (BIC), which is one method of optimizing both goodness of fit and parsimony; and (3) A set of factor retention criteria that have been used in Weisman et al.’s (2017) previous work, in which they began with the maximal number of factors according to an analysis of degrees of freedom, and retained factors that met all three of the following criteria: (a) had eigenvalues greater than 1.00, (b) individually accounted for greater than 5% of the shared variance before rotation, and (c) were the “dominant” factor (the factor with the strongest absolute factor loading) for at least 1 mental capacity after rotation. For each study, our interpretation of how best to characterize the dataset (i.e., how many factors we observe) was determined by the degree of consensus among these three protocols and the interpretability of the retained factors under each protocol.

Here we focus our interpretations on varimax-rotated solutions, which constrain all factors to be orthogonal. (See OSM for solutions using oblique [“oblimin”] rotations, which allow for correlated factors.)

## Secondary analysis: Regression analyses of factor scores

Having inferred a conceptual structure for a given group of participants via EFA, we then sought to examine attributions of mental capacities to the particular target entities included in each study within this conceptual structure: To what extent did participants attribute each of the fundamental components of mental life revealed by EFA to a given target entity, and how did this attributions vary with age (either within an age group or between age groups)? To explore this question, we examined “factor scores”—summaries of each participant’s attributions of each of factors revealed by EFA. We used the correlation-preserving “ten Berge” method (as implemented in the “psych” package; Revelle, 2018), omitting participants with any missing data. This yielded one factor score for each of the factors revealed by EFA for each participant.

We analyzed these factor scores via mixed effects Bayesian regression analyses using the “brms” package for R (Bürkner, 2017). In all of these analyses, categorical variables were effect-coded, continuous variables were centered at the mean, and we included the maximal random effect structures given the design for the relevant study. Further details varied by study, depending on the number of target entities included in that study, the number of factors revealed by EFA for the relevant group(s) of participants, and the goals of the analysis (e.g., comparing two age groups vs. examining continuous effects of age within one or more groups of participants).

# Study 1

The goal of Study 1 was to develop a child-friendly version of Weisman et al.’s (2017) study paradigm and conduct an initial exploration of this conceptual structure among children.

Pilot testing suggested that children as young as 7 years of age found the paradigm easy and enjoyable, and work on the development of lay biology and psychology has suggested that these concepts may continue to develop well into middle childhood (e.g., Carey, 1985; Hatano & Inagaki, 1997; Piaget, 1929; cf. Gelman & Opfer, 2002). Thus, we targeted 7- to 9-year-old children for our first child sample. We also recruited a group of adults to validate our child- friendly paradigm, i.e., to evaluate whether it replicated Weisman et al.’s (2017) original work with adults. In Weisman et al.’s (2017) original studies, participants evaluated a target character on 40 mental capacities using a seven-point Likert-type scale. Pilot testing suggested two necessary modifications for children: rewording some of the mental capacity items, and using a simpler, three-point response scale (no, kinda, or yes; see “Overview of Methods and Analyses, above”).

## Method

### Participants

400 people participated in this study.

Adults (*n*=200) participated via Amazon Mechanical Turk (MTurk) in July 2016. Adult participants had gained approval for at least 95% of their previous work on MTurk; had MTurk accounts based in the US; and indicated that they were at least 18 years old. Adults were paid $0.30 for approximately 2-3 minutes of their time (median duration: 2.48 min).

According to self report, the adult sample ranged in age from 18-65 years (median: 31y) and was roughly split between women (47%) and men (52%; 1% of participants identified as some other gender or opted not to disclose). Adults predominantly identified as White (81%; 7.5% identified as more than one race/ethnicity, and 3.5% as any other race/ethnicity). The vast majority of adults reported English being their only native language (91%; an additional 7% indicated that English was one of multiple native languages for them.)

Children (*n*=200) participated at one of several San Francisco Bay Area museums or at their younger sibling’s preschool between July-December 2016. The study took most children under 10 minutes to complete (median duration: 5.18 min). An additional 12 children participated but were excluded for being outside the target age range (*n*=7), being of unknown age (*n*=4), or being shown a target character other than a beetle or a robot (*n*=1). Children received a small thank-you gift (e.g., a sticker) for their participation.

Children ranged in age from 7.01-9.99 years (median: 8.31y). According to parental report, the child sample included slightly more girls (56%) than boys (42.5%; 1.5% of children’s gender was non-binary or unknown). Parents predominantly identified their children as White (31.5%), multiracial (20.5%), East Asian (12%), or South Asian (10%; 2.5% of chidlren were identified as any other race/ethnicity, and 22.5% of children’s race/ethnicity was unknown). Roughly half of parents (46.5%) reported that their child was bilingual (though, anecdotally, parents’ interpretations of “bilingual” ranged from taking classes at school to speaking a langauge at home).

### Materials and procedure

Following Weisman et al. (2017, Studies 1-3), participants were randomly assigned to assess the mental capacities of one of two “edge cases” in social reasoning: a beetle (*n*=98 adults, *n*=104 children) or a robot (*n*=102 adults, *n*=96 children). Because beetles are animals and robots are artifacts, this pair provides insight into the role of biological life in attributions of mental life—an issue of particular interest from a developmental perspective, given the long history of work on the development of the animate-inanimate distinction and its relation to folk psychology. Most critically for our bottom-up approach to uncovering conceptual structure, the “mental lives” of these entities are controversial: People differ in their assessments of the mental capacities of beetles and robots. This allowed us to address the following question: When participants disagree about the mental capacities of some entity, which capacities “go together”?

Instructions to participants focused on the idea that we wanted to know what participants thought “[beetles/robots] can do and can not do.” Participants rated the target character on 40 mental capacities, presented in a random order for each participant. On each trial, participants responded *no*, *kinda*, or *yes* to the question “Do you think a [beetle/robot] can…?” The three response options were visible throughout the experiment.

The 40 mental capacities were designed to be as close as possible to those in Weisman et al.’s (2017) original studies, while being comprehensible to children in early elementary school. This set of items included physiological sensations related to biological needs (e.g., *get hungry*); emotional experiences (e.g., *feel happy*); perceptual abilities (e.g., *hear sounds*); cognitive abilities (e.g., *remember things*); capacities related to autonomy or agency (e.g., *decide what to do*); social abilities (e.g., *feel guilty*); and several additional items (e.g., *be aware of itself*). Each of these a priori categories included at least five items of varying valence, complexity, and phrasing. (See Figure 1 for the full list of mental capacities included in Study 1.)

We also prepared a short definition for each item, so as to be consistent in our responses to participants (particularly children) if they asked for clarification. Children were encouraged at the beginning of the study to ask questions if they did not know what a word meant, in which case they given these definitions; adults were told that they could access these definitions by hovering over the text on the computer screen. Pilot testing suggested that seven items required clarification for most children, so these items were always accompanied by their definitions from the beginning of the trial (for both adults and children), as follows: *have a personality, like when someone is shy and somebody else is silly*; *have beliefs, like when you think something is true*; *feel pleasure, like when something feels really good*; *have desires, like when you really want something*; *have self- control, like when you stop yourself from doing something you shouldn’t do*; *have goals, like when you’re trying hard to do something or make something happen*; and *feel sick, like when you feel like you might throw up*.

Adults completed the study by clicking through a website at their own pace, with one trial presented on each page and no ability to go backwards. Children completed the study on an experimenter’s laptop computer. The experimenter read the instructions and the first several trials out loud, requesting verbal responses from the child and selected his or her response for her; after several trials, the experimenter gave the child the option to continue independently (reading the questions and selecting their answers themselves) if they desired. Roughly half of participants completed the remainder of the task independently.

### Data processing

We dropped trials with response times that were faster than a preset criterion of 250ms (*n*=3 child trials, *n*=97 adult trials) and retained participants regardless of skipped trials (*n*=55 child trials, *n*=1 adult trial). Overall, only 1% of adult trials and 1% of child trials were missing data.

## Results

### Conceptual structure: Adults

Two or our three methods for determining how many factors to retain (minimizing BIC and Weisman et al.’s preset factor retention criteria) suggested retaining three factors. (Parallel analysis suggested retaining 4 factors; see SOM.)

After rotation, the first factor corresponded primarily to capacities for self- and other-relevant emotions—a suite of capacities that we (following Weisman et al.) will refer to as *HEART*. It was the dominant factor for such items as *feel happy*, *feel joy*, *feel proud*, and *feel sad*, and accounted for 37% of the shared variance in the rotated three-factor solution.

The second factor corresponded primarily to physiological sensations related to biological needs—a suite of capacities that we (following Weisman et al.) will refer to as *BODY*. It was the dominant factor for such items as *feel pain*, *feel scared*, *feel tired*, and *get hungry*, and accounted for 37% of the shared variance in the rotated three-factor solution.

The third factor corresponded primarily to perceptual-cognitive abilities to detect and use information about the environment—a suite of capacities that we (following Weisman et al.) will refer to as *MIND*. It was the dominant factor for such items as *figure out how to do things*, *make choices*, *recognize somebody else*, and *sense whether something is close by or far away*, and accounted for 26% of the shared variance in the rotated three-factor solution.

See Figure 1 for all factor loadings. (Note that for the sake of consistency across studies and comparison with Weisman et al.’s work, we have plotted these factors in the same order for all studies: *BODY*, *HEART*, and *MIND*.)

In sum, as Weisman et al.‘s original studies, a three-factor structure emerged from adults’ mental capacity attributions, characterized by a distinction between body, heart, and mind. This suggests that our child-friendly paradigm was valid: Using reworded items and a three-point response scale elicited the same intuitive ontology of mental life, among adults, that has been revealed by more complex, “adult-friendly” experimental paradigms.

### Conceptual structure: Children (7-9y)

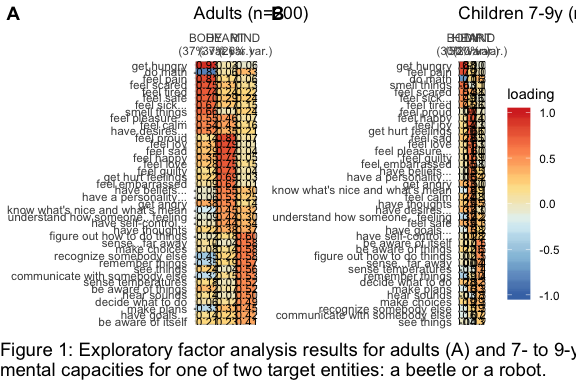
Again, our three methods for determining how many factors to retain all suggested retaining three factors.

After rotation, the first factor corresponded primarily to social-emotional abilities. An analysis of factor congruence confirmed that this factor was most similar to adults’ *HEART* factor (cosine similarity with *HEART*: 0.97; with *BODY*: 0.41; with *MIND*: 0.43). It was the dominant factor for such items as *feel happy*, *feel joy*, *feel proud*, and *get hurt feelings*, and accounted for 50% of the shared variance in the rotated three-factor solution.

The second factor corresponded primarily to physiological sensations. An analysis of factor congruence confirmed that this factor was most similar to adults’ *BODY* factor (cosine similarity with *BODY*: 0.91; with *HEART*: 0.27; with *MIND*: 0.02). It was the dominant factor for such items as *feel pain*, *feel scared*, *get hungry*, and *smell things*, and accounted for 30% of the shared variance in the rotated three-factor solution.

The third factor corresponded primarily to perceptual-cognitive abilities. An analysis of factor congruence confirmed that this factor was most similar to adults’ *MIND* factor (cosine similarity with *MIND*: 0.94; with *HEART*: 0.35; with *BODY*: 0.02). It was the dominant factor for such items as *be aware of itself*, *be aware of things*, *figure out how to do things*, and *sense whether something is close by or far away*, and accounted for 20% of the shared variance in the rotated three-factor solution. (See Figure 1 for all factor loadings.)

In sum, like adults, children’s mental capacity attributions were dominated by a three-way distinction between physiological, social-emotional, and perceptual-cognitive abilities—i.e., body, heart, and mind.



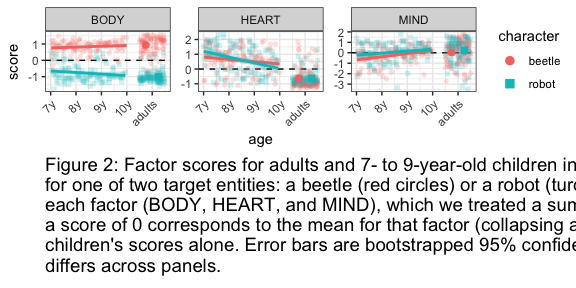
### Application of the concept: Children’s vs. adults’ attributions of mental life

To formally compare responses from children and adults, we considered the full, combined dataset and examined factor scores by age group.

Two or our three methods for determining how many factors to retain (minimizing BIC and Weisman et al.’s preset factor retention criteria) suggested retaining three factors. (Parallel analysis suggested retaining 4 factors; see SOM.)

EFA using the combined dataset revealed three factors that met our retention criteria. Unsurprisingly, these three factors were very similar to those revealed for adults and children analyzed independently: They corresponded to social-emotional abilities (*HEART*), physiological sensations (*BODY*), and perceptual-cognitive abilities (*MIND*; see SOM).

The purpose of this combined EFA was to examine differences in adults’ and children’s evaluations of beetles and robots within this three-part structure. To do so, we derived factor scores (via the ten Berge method) using the rotated three-factor solution (using medians by capacity to impute missing values). This yielded three scores for each participant, corresponding, in principle, to holistic judgments of the social-emotional, physiological, and perceptual-cognitive abilities of the target character the participant evaluated. (Note that each of these three scores takes into account factor loadings for all 40 mental capacities, as listed in SOM, Table S1.) This allowed us to examine the effects of age group (adult, child), character (beetle, robot), and factor (*BODY*, *HEART*, *MIND*) on these scores via mixed effects linear regression. See Table XX for the results of a maximal model and Figure 2 for scores by age group, age (for children), factor, and character.



**BOOKMARK**

Collapsing across age groups and factors, factor scores suggest that participants generally attributed fewer mental capacities to the robot than the beetle (b=XX). However, this appears to be entirely due to the huge discrepancy between characters in the physiological domain; the difference between characters was reduced to nothing in the social-emotional domain (b=XX), and reversed in the perceptual-cognitive domain (b=XX). Collapsing across entities (beetle, robot), children tended to attribute more mental capacities adults (b=XX), but this was driven primarily by the social-emotional domain (b=XX), and was reversed in the perceptual-cognitive domain (b=XX).

Scores in the physiological and perceptual-cognitive domains were very similar for children and adults: Both children and adults marked a clear difference between the robot and the beetle in the physiological domain (Figure XX, center), in line with the animate–inanimate distinction; and both age groups credited the robot with slightly greater perceptual-cognitive skills than the beetle (right). In contrast, in the social-emotional domain (Figure XX, left) both the beetle and the robot received rather low scores among adults, but very high scores among children. See Figure XX for raw counts of no, kinda, and yes responses for all items, grouped by character, age group, and dominant factor.

In sum, we observed only minor differences between children and adults in their attributions of physiological and perceptual-cognitive abilities to beetles and robots—but a major difference in the social-emotional domain: Relative to adults, children tended to credit both beetles and robots with much greater social-emotional abilities.

## Discussion

XX

A number of additional or alternative latent factors could have emerged from the factor analysis of children’s responses. For example, children might have distinguished primarily between internal experience and external action (Gray et al., 2007), or they might have demonstrated finer-grained groupings of mental capacities based on phrasing, rote knowledge, etc. Instead, the latent conceptual structure underlying children’s responses appears to be very similar to that of adults.

# Study 2

The goals of Study 2 were twofold. First, we aimed to (conceptually) replicate and extend our findings with 7- to 9-year-old children in Study 1. To this end, we decided to extend the target entities included in Study 2 to include not just the two “edge cases” from Study 1 (a beetle and a robot), but a wider range of animate beings (a bird, a goat, and an elephant) and inanimate objects (a computer, a teddy bear, and a doll). In Weisman et al.’s original studies, this approach yielded a highly similar conceptual structure to the “edge case” approach (Weisman et al., 2017; Study 4). We reasoned that if this three-part conceptual structure is stable and robust by the age of 7- to 9-year-olds, it should manifest among 7- to 9-year-old children under the same range of conditions that elicit this structure from adults.

Second, we aimed to assess the earlier development of this conceptual structure in a group of younger children. We targeted 4- to 6-year-old children for our younger age group because this has been identified as a period of dramatic cognitive and social development in many domains. Many studies have documented shifts in children’s abilities to take others’ perspectives, represent false beliefs, and integrate representations of intentions and outcomes in evaluating moral responsibility (for reveiws, see Flavell, 1999; Wellman, 2015). The preschool years have also been the focus of a rich tradition of work on lay biology and the animate-inanimate distinction extending back nearly a century (e.g., Carey, 1985; R. Gelman, Spelke, & Meck, 1983; Medin, Waxman, Woodring, & Washinawatok, 2010; Piaget, 1929; for a review, see S. Gelman & Opfer, 2002). All of these accounts make the case that becoming a sophisticated reasoner—and particularly a sophisticated social reasoner—requires substantial refinement of one’s representations of others’ experiences, beliefs, desires, and needs. Might these refinements include shifts in children’s intuitions about the fundamental components of mental life?

Pilot testing suggested that this would require making a briefer experimental paradigm with fewer than the 40 questions included in Study 1; limiting the list to 20 questions seemed to allow children as young as 4 years of age to complete the study easily and without getting bored or frustrated, while still including enough items to enable us to proceed with our exploratory “dimensionality reduction” approach to uncovering conceptual structure. To validate this modified paradigm, we also included a sample of adults.

## Method

### Participants

XX people participated in this study.

Adults (n=XX) participated via MTurk in September 2018. Adult participants had gained approval for at least 95% of their previous work on MTurk; had MTurk accounts based in the US; and indicated that they were at least 18 years old. Adults were paid $0.45 for approximately XX minutes of their time (median duration: XX min). An additional XX adults participated but were excluded for failing to pass one or more attention checks (e.g., “Please select no”; n=XX) or for failing to respond sensibly to an open-ended question about what they had been asked to do in the study (e.g., copying and pasting text from the question, writing “good study,” or describing a different study, e.g., “I wrote an essay about nature”; n=XX). Among the final sample of XX adults, XX% identified as women and XX% as men. Adults predominantly identified as White (XX%; XX% as Black, XX% as Asian; XX% as XX), and reported being between XX and XX years of age (median age: XX years). XX OTHER DEMOGRAPHICS: Native language, education, religion as a child, religion now, job.

Data collection with children occurred between January-June 2017. Our planned sample size was 120 older and 120 younger children, but we also retained a handful of extra participants who completed the study on the final day of data collection for each group.

The group that we will refer to as “older children” (n=123) ranged in age from 7.09-9.99y (median: 8.57y), and participated at museums in the San Francisco Bay Area; the median study duration for this group was 2.70min. According to parental report, XX children (XX%) were girls and XX (XX%) were boys; for XX children (XX) gender was either non-binary or unknown. Parents predominantly identified their children as White (XX%) or Asian (XX%; XX% as Black, XX% as Asian; XX% as XX). XX OTHER DEMOGRAPHICS: Language.

“Younger children” (n=124) ranged in age from 4.00-6.99y (median: 5.03y), and participated either at their preschool (68%) or at a museum (32%); the median study duration for this group was 3.58min. According to school records, XX children (XX%) were girls and XX (XX%) were boys; for XX children (XX) gender was either non-binary or unknown. Parents predominantly identified their children as White (XX%) or Asian (XX%; XX% as Black, XX% as Asian; XX% as XX). XX OTHER DEMOGRAPHICS.

An additional 7 children participated but were excluded for being outside the target age ranges.

We grouped children into two age groups because our primary planned analysis (EFA) is a group-level analysis of the consensual conceptual structure, and is not designed to model continuous participant-level variables like exact age. Our goal in this study was to examine discrete “snapshots” of this conceptual structure at two points in this continuous developmental trajectory.

### Materials and procedure

Participants were assigned to evaluate one of the following target entities: an elephant, a goat, a mouse, a bird, a beetle, a teddy bear, a doll, a robot, or a computer (n=XX-XX per character, per age group). This allowed us to address the following question: When target entities are perceived to vary in their mental capacity profiles, which capacities “go together”?

Participants were assigned to target characters randomly, with two exceptions: The doll and teddy bear conditions were run last for older children (but included in the initial randomization scheme for adults and younger children); and toward the end of data collection children were assigned to conditions that had the fewest participants. As in Study 1, a vivid, high-resolution photo of the target character in a naturalistic context was visible for the duration of the study.

Instructions were identical to Study 1. Participants rated the target character on 20 mental capacities, presented in a random order for each participant. On each trial, participants responded *no*, *kinda*, or *yes* to the question “Do you think a [beetle] can…?” As in Study 1, adults completed the study by clicking through a website at their own pace, with one trial presented on each page and no ability to go backwards, and children completed the study on an experimenter’s laptop computer. Following Study 1, for older children the experimenter read the instructions and the first several trials out loud, requesting verbal responses from the child and selected his or her response for her; after several trials, the experimenter gave the child the option to continue independently if they desired. All younger children heard all questions read aloud by the experimenter and responded verbally.

The 20 mental capacities were a subset of the 40 items used in Study 1, chosen to include physiological sensations, emotional experiences, perceptual abilities, cognitive skills, capacities related to autonomy or agency, and social abilities; and to include some of the strongest-loading items for each of the three factors uncovered in Study 1 (see Figure XX). As in Study 1, each mental capacity was associated with a short, preset definition. Children were encouraged at the beginning of the study to ask questions if they did not know what a word meant, in which case they were given these definitions. (Adults did not have access to these definitions.)

### Data processing

We planned to drop trials with response times that were faster than a preset criterion of 250ms, but there were none. As in Study 1, we retained participants regardless of skipped trials (n=0 trials among older children, 30 trials among younger children). Overall, none of adults or older children’s trials, and only 1.21% of younger children’s trials, were missing data.

## Results

XX

## Discussion

XX

# Study 3

The primary goal of Study 3 was to (conceptually) replicate and extend our findings with 4- to 6-year-old children in Study 2. In light of the concerns about vocabulary, attention, and use of the response scale among younger children in Study 2, we designed an even more child-friendly version well suited for young preschoolers, streamlining the experimental protocol, providing more scaffolding for the response scale, and including only vocabulary items that were pre-tested to be familiar to young preschool children (see “Method”).

A second goal of Study 3 was to assess younger children’s attributions of mental life to edge cases. Recall that, in Study 1, older children attributed more social-emotional abilities (“heart”) to both beetles and robots than did adults, despite sharing adults’ three-part conceptual structure; this “edge case” approach of Study 1 allowed us to estimate older children’s aggregate attributions to these entities with precision and to chart changes in these attributions over age continuously within the range of 7-9 years, because large numbers of children assessed each of the two target entities (rather than participants being distributed across a wider range of target entities). In Study 3, we returned to this approach, limiting the target characters to a beetle and a robot, so as to follow up on this secondary finding with a younger group of children.

Finally, Study 3 was designed in anticipation of developing participant-level analyses to supplement the group-level EFAs discussed so far (see “Continuous development at the participant level: A re-analysis of Studies 1-3,” below). With this general goal in mind, we asked each child to assess the mental capacities of both a beetle and a robot, so as to maximize the number of observations per participant for these anticipated participant-level analyses.

## Method

### Participants

XX people participated in this study.

Adults (n=XX) participated via Amazon Mechanical Turk (MTurk) in September 2018. Adult participants had gained approval for at least 95% of their previous work on MTurk; had MTurk accounts based in the US; and indicated that they were at least 18 years old. Adults were paid $0.75 for approximately XX minutes of their time (median duration: XX min). An additional XX adults participated but were excluded for failing to pass one or more attention checks (e.g., “Please select no”; n=XX) or for failing to respond sensibly to an open-ended question about what they had been asked to do in the study (see Study 3 for examples; n=XX).

Among the final sample of XX adults, XX% identified as women and XX% as men. Adults predominantly identified as White (XX%; XX% as Black, XX% as Asian; XX% as XX), and reported being between XX and XX years of age (median age: XX years). XX OTHER DEMOGRAPHICS: Native language, education, religion as a child, religion now, job. XX children participated in this study, which was conducted between January 2018 and XX 2018. Our planned sample size was 100 children, but we also retained a handful of extra participants who completed the study on the final days of data collection. Children ranged in age from 4.00-XXy (median: XXy), and participated at their preschool; exact duration was not recorded, but the full study session generally took less than 10 minutes. According to school records, XX children (XX%) were girls and XX (XX%) were boys; for XX children (XX) gender was either non-binary or unknown. Parents predominantly identified their children as White (XX%) or Asian (XX%; XX% as Black, XX% as Asian; XX% as XX). XX OTHER DEMOGRAPHICS.

An additional XX children participated but were excluded for being outside the target age range.

### Materials and procedure

Materials and procedure were adapted to be more appropriate for young preschoolers, with two primary goals in mind: Streamlining the experimental protocol to improve children’s comprehension and attention to the task, and limiting mental capacities to words that are highly familiar to young preschool children.

In order to streamline the experimental protocol, we moved the task off of the computer (for children but not adults), instead using printed photographs of the target entities (measuring approximately 5 x 8 inches, printed in color and laminated) and recording children’s responses by hand (rather than using the computer). At the time of testing, the experimenter and child sat side by side, with the photograph placed directly in front of the child for the duration of the task.

We also streamlined the introduction to the task. The experimenter began by placing the photograph of the first target entity in front of the child and asking, “Can you tell me what this is?” If a child provided an answer other than “beetle” or “robot,” the experimenter said, “I’m going to call it a [beetle/robot].” Then the experimenter said, “We’re going to play a game about [beetles/robots],” reminded children, “if you ever want to stop playing, you can just let me know and we’ll go back to the classroom” (per preschool protocol), and then launched into the first question (e.g., “Can beetles get sad?”).

We scaffolded children’s use of the three-point response scale by providing a physical representation of the scale, consisting of three large boxes, separated by blank space, containing the words NO, KINDA, and YES written in large font with all capital letters (to aid children with at least some reading skills in recognizing these words); color-coded according to the intensity of response (NO = very light blue, KINDA = medium blue, YES = dark blue); and ordered from left (NO) to right (YES). Each box measured approximately 2 x 4 inches; the boxes were laminated with slightly less than 1 inch of empty space between them (through which the table was visible). In addition to providing these visual and spatial cues to the fact that there were three response options—no, yes, and something conceptually and literally “in between” these extremes—we also added a line to the experimenter’s script describing (and then reiterating) these response options on the first three trials (“You can say no [pointing to NO], kinda [pointing to KINDA], or yes [pointing to YES]”). The experimenter repeated these options on the first three trials for all children, and on any other trials where a child took more than a few seconds to answer or provided a response other than saying “yes,” “kinda,” or “no” or clearly pointing to one of these options on the response scale.

For each target entity, children answered 18 questions about its mental capacities. These questions were designed to include six clear examples of bodily, social-emotional, and perceptual-cognitive abilities, according to Studies 1-2 and Weisman et al.’s (2017) original studies with US adults. Bodily abilities included feel hungry, get thirsty, feel sick, feel tired, get scared, and smell things. Social-emotional abilities included love someone, hate someone, feel happy, get sad, feel sorry, get lonely. Perceptual-cognitive abilities included see, hear, think, remember things, know stuff, figure things out. These items were chosen to as short as possible and highly familiar to young preschool children; they were selected from a larger pilot study in which we asked 3- to 5-year-old children to complete stories that began with each of these mental capacities as a premise (e.g., “Let’s imagine a person who loves someone. What happens next?”; “Now let’s pretend that someone remembers something. What happens next?”) and judged the appropriateness of their story completion. We further designed the set of items so that each category included a variety of phrasings (e.g., feel hungry vs. get thirsty; remember things vs. know stuff) and valences when appropriate (e.g., feel happy vs. get sad); when possible, we aimed to have phrasings vary orthogonally with categories, such that “get” and “feel” appeared roughly equally often among the bodily and social-emotional items, and “things” appeared equally often among the bodily and perceptual-cognitive items. As in Studies 1-2, each mental capacity was associated with a short, preset definition, which was read to children if they expressed uncertainty about what a word meant or did not respond after prompting use of the response scale.

Children first assessed all 18 mental capacities for one of the target characters (e.g., the beetle), then completed an easy jigsaw puzzle featuring clothing and accessories appropriate for a rainy day (which took about 30-60 second to complete), and finally assessed all 18 mental capacities for the other target character (e.g., the robot).

This modified procedure—particularly moving the experiment off of the computer for children—required changes to the experimental design. Rather than randomly assigning children to assess the beetle first or the robot first, the order of target entities was counterbalanced in advance. Likewise, rather than asking about the 18 mental capacities in a random order, questions about the first target entity were asked in one of 8 pre-made random orders, and questions about the second target entity were asked in the reverse order. The order of the target entities (beetle-robot or robot-beetle) and the order of the mental capacity questions (1-8) were fully crossed across participants.

Adults participated in an online version of this same task, without a break between target characters. As in Studies 1 and 2, adults clicked through a website at their own pace, with one trial presented on each page and no ability to go backwards.

### Data processing

We did not record response times or use this as a criterion for inclusion. As in Studies 1-2, we retained participants regardless of skipped trials (n=XX trials; XX% of all trials).

## Results

XX

## Discussion

XX

# Assessing age-related changes in conceptual structure continuously: A re-analysis of Studies 1-3

Taken together, exploratory factor analyses (EFAs) of different age groups suggest that that conceptual structure underlying 4- to 6-year-old children’s mental capacity attributions (Studies 2-3) differs from the conceptual structure underlying 7- to 9-year-old children’s and adults’ mental capacity attributions (Studies 1-2). This data-driven, “bottom-up” approach to inferring conceptual structure at the group level provides novel insights into how children’s representations of mental life might evolve across childhood that we would not have anticipated a priori. A similar pattern of differences between younger and older children emerged organically across multiple studies (Study 2 and Studies 1 vs. 3), giving us reason to think that they reflect something real about children’s developing understanding of mental life.

However, this approach to analyzing data comes at some cost. EFA is a group-level analysis, capable of revealing latent structures underlying covariance patterns across a large set of observations; using this approach to study conceptual structure at the individual level would require hundreds of observations per participant, which is not feasible with young children. Even collecting enough observations per age group was no small feat; in order to facilitate data collection for Studies 1-3 we opted for fairly wide age ranges spanning 3 years for each group (4-6 years and 7-9 years). Categorizing children as “younger” vs. “older” using such wide age ranges has provided intriguing snap-shots of children at different points in development—but this is, of course, a rather crude way of characterizing developmental change.

Between Studies 1-3, however, a total of 545 children between the ages of 4.00-9.99 years each provided between 20-40 responses to questions about various target entities’ mental capacities. With a different kind of analysis, this dataset has the potential to address questions about conceptual development at a much more fine-grained level than what group-level EFAs have revealed so far. With this in mind, we now present a novel re-analysis of Studies 1-3, aimed at characterizing age-related differences in conceptual structure continuously across development. Our goal in these post-hoc analyses was to assess the differentiation of what we’ve called body, heart, and mind in individual children—a kind of non-parametric, participant- level analysis meant to parallel the EFAs reported above.

We based this analysis on the intuition that a child who differentiates clearly between two categories of mental capacities (e.g., physiological sensations of the body vs. social-emotional abilities of the heart) will evaluate mental capacities related to these categories somewhat independently. Such a child will sometimes end up endorsing mental capacities in one category while rejecting mental capacities in the other (e.g., endorsing most physiological sensations but rejecting most social-emotional abilities)—whereas a child who does not differentiate between these categories might be more likely to endorse or reject across the board (e.g., endorsing equal numbers of physiological and social-emotional items). Of course, depending on the target character they happen to evaluate, even children with clearly differentiated categories might end up endorsing equal numbers of capacities in both. But if the differentiation of two categories becomes stronger over development, we might expect that, on average, the difference in the number of endorsements between these categories would increase with age.

Following this logic, we re-analyzed the data from Studies 1-3 to assess changes in conceptual structure continuously between 4-9 years of age.

## Method

### Participants

We included all of the child participants from Studies 1-3 in this analysis (total N=545 children). This combined sample of children ranged in age from 4.00-9.99 years (median: XX years); see Methods for Studies 1-3 for further demographic information.

### Data preparation

Following the logic outlined above, we designated a set of mental capacity items to represent categories of body, heart, and mind for each study and tallied up the number of “endorsements” (responses of yes or kinda) within each category for each child (separating out children’s attributions to the beetle vs. the robot, for Study 3). For the sake of comparability across studies and across categories, we included exactly six mental capacities in each category.

For Study 3, these categories had already been defined in the initial design of the study: To reiterate, the body category included feel hungry, get thirsty, feel sick, feel tired, get scared, and smell things; the heart category included love someone, hate someone, feel happy, get sad, feel sorry, get lonely; and the mind category included see, hear, think, remember things, know stuff, figure things out.

For Studies 1-2, we used adults’ EFA results from Study 1 to define categories. For the body category, we included get hungry, feel pain, smell things, feel scared, feel sick, and feel tired. These six items were among the seven strongest positive factor loadings on Factor XX among adults in Study 1 (all loadings ≥0.66), were included in both Studies 1 and 2, and had maximal overlap with the body category as defined for Study 3 (five of six items were nearly identical, with the exception of feel pain instead of get thirsty).

For the heart category, we included feel love, feel happy, feel sad, get angry, feel guilty, and get hurt feelings. These six items were among the ten strongest positive factor loadings on Factor XX among adults in Study 1 (all factor loadings ≥0.51), were included in both Studies 1 and 2, and had maximal overlap with the heart category as defined for Study 3 (three of six items were nearly identical, and the others were matched in valence and general meaning: get angry instead of hate someone; feel guilty instead of feel sorry; and get hurt feelings instead of get lonely).

Finally, for the mind category, we included figure out how to do things, remember things, be aware of things, sense whether something is close by or far away, sense temperatures, and make choices. These six items were among the ten strongest positive factor loadings on Factor XX among adults in Study 1 (all factor loadings ≥0.50) and were included in both Studies 1 and 2. They also had relatively strong overlap with the mind category as defined for Study 3: Two of six items were nearly identical, and two others were matched in general meaning (be aware of things instead of see; sense temperatures instead of hear); the others (sense whether something is close by or far away and make choices) were not particularly well-matched to the Study 3 substitutes (know stuff and think), but allowed us to use the same set of items to define the mind category for Studies 1 and 2, which we decided was our priority.

We note that there are many “researcher degrees of freedom” in these choices; we have striven to be as transparent as possible about our process, and encourage the curious reader to explore other choices using our data (openly available at XX).

This left us with a dataset in which each child was associated with “endorsement tallies” between 0-6 for body, for heart, and for mind, for each of the target entities that child assessed (Studies 1-2: one target per child; Study 3: two targets per child); this data also included each child’s exact age.

### Analysis plan

Following the logic described above, we compared differences in children’s endorsement tallies across pairs of categories: body minus heart, body minus mind, and heart minus mind, interpreting large differences between categories as reflecting strong differentiation of those two categories and age-related increases in differences to reflect increasing differentiation over development.

[OPTION #1: We conducted three separate mixed-effects Bayesian regression analyses—one for each pair of categories—to analyze these age-related changes in the absolute value of differences (theoretical range: 0-6). For each analysis, we regressed children’s exact age (centered at the mean, XX years) onto the difference between the two categories in question (e.g., body minus heart), using Poisson distributions and maximal random effects structures. XX ALSO RAW DIFFERENCES?]

[OPTION #2: We conducted a mixed-effects Bayesian regression analysis to analyze differences in endorsements across categories and whether they varied with children’s age, using children’s exact age (centered at the mean, XX years), the category (coded with orthogonal contrasts comparing body and heart to mind, and body to heart), and an interaction between age and category to predict the endorsement tally (theoretical range: 0-6). We used Poisson distributions to model these tallies, and maximal random effects structures.]

## Results

XX

## Discussion

XX

# General Discussion

## Summary

The current studies suggest that children’s conceptual representations of mental life undergo substantial development between the ages of 4-9 years—in two distinct senses.

First, analyses of the correlational structure of children’s mental capacity attributions (EFAs) suggested changes in conceptual structure: Between early childhood (4-6 years; Studies 2-3) and middle childhood (7-9 years; Studies 1-2), children’s sense of which mental capacities “go together” seemed to unfold into a higher-dimensional space, characterized by stronger differentiation of the social-emotional abilities that we’ve called “heart” from the physiological sensations of the body and the perceptual-cognitive abilities of the mind. Meanwhile, analyses of factor scores—summaries of which mental capacities children tended to attribute to which target entities—revealed a second kind of age-related difference: what we might call changes in the application of the concept. These differences were clearest in Studies 1 and 3, which focused on two “edge cases” in social reasoning (beetles and robots): Between the ages of 4-6 years (Study 3), XX FILL IN, while between the ages of 7-9 years (Study 1), attributions of body to beetles and robots were relatively stable and adult-like, attributions of heart to these “edge cases” started out higher than adults and decreased to nearly adult-like levels by the age of 9 years, and attributions of mind started out lower than adults and increased to nearly adult-like levels by the age of 9 years.

XX GD