Précis of *The Development of Conceptual Representations of Mental Life*

Kara Weisman

Ph.D. in Psychology, Stanford University, 2019

Overview

Attributions of thoughts, beliefs, desires, intentions, emotions, perceptions, and sensations are at the core of human social life—but the experiences and abilities that make up our "mental lives" vary along many dimensions. Some are closely related to specific bodily organs (e.g., hearing, hunger), and others less obviously so (sadness, belief). Some are positively or negatively valenced (e.g., happiness, pain), others are more neutral or variable in valence (memory, smell). Some mental capacities involve taking in information about the environment (e.g., vision, taste), while others involve storing or updating that information (memory, learning), or using it to bring about changes in the world (planning, choice). In a given social context, certain mental states might be valued more than others (e.g., anger vs. guilt, excitement vs. contentment); and people might believe that certain mental capacities are shared by a wide range of entities, while others are limited to a smaller subset of beings (e.g., fear vs. pride). This makes conceptual representations of mental life a fascinating case study of abstract reasoning and its development: How do people come to represent this complex conceptual space?

In this dissertation, I describe a series of studies designed to explore such conceptual representations of mental life among US children and adults. My approach unites rich traditions of research in cognitive and developmental psychology on the animate-inanimate distinction, lay biology and psychology, and theory of mind with intuitions from ancient philosophy, methods from social psychology, and modern, "bottom-up" statistical analyses. In these studies, I use participants' answers to simple questions—e.g., "Can a beetle get hungry?", "Can a goat feel love?", "Can a robot remember things?"—to address three key questions about representations of mental life:

1) What are the conceptual units that anchor representations of mental life?
2) What is the structural organization governing the relationships among these conceptual units?
3) How do people deploy their conceptual representations of mental life to reason about specific entities in the world—namely, animate beings vs. inanimate objects?

The dissertation includes seven large-scale studies with adults (total N=1862) and three studies of 4- to 9-year-old children (N=493). I apply various exploratory analysis techniques to each of these datasets in order to investigate age-related changes in these three key aspects of conceptual representations of mental life. These findings give rise to a new theory of conceptual development in this domain, with implications for social reasoning in particular, and for conceptual development more broadly.

Theoretical Foundations

Seminal findings in cognitive and developmental psychology. Many classic traditions of work within cognitive and developmental psychology have touched on children's understanding of mental life—converging to suggest that the period of development between roughly 4-10 years of age is a time of rapid change in this conceptual domain.

Early explorations of children's understanding of mental life were rooted in a more basic question: How do children understand "life" at all? Capacities for desires, goals, intentions, and other mental states are one of the key features that distinguish animate beings like humans and animals from inanimate objects (Gelman & Spelke, 1981). Indeed, beginning with Piaget's (1929) classic work on what he termed "childhood animism," attributions of such mental states to inanimate beings have been considered a hallmark of an immature understanding of life and animacy. Since then, many foundational figures in the field of cognitive development—Carey, Flavell, R. Gelman, S. Gelman, Keil, Medin, Spelke, Waxman, and Wellman, among others—have taken up the questions of whether, when, and why young children might attribute mental states to inanimate objects, and what might drive the decline of these "over-attributions" over development.

A pivotal moment in this tradition was the publication of Carey's (1985) exploration of conceptual change in the domains of lay psychology and biology. Carey documented dramatic changes in children's understanding of what it means for something to be alive; why living do things like eat and sleep; and which set of entities in the world have these properties. Prior to roughly 7 years of age, children in Carey's studies appeared to base their assessment of whether entities are alive on judgments of their similarity to humans, and to ground their reasoning about activities like eating and sleeping in their understanding of human social-psychological behavior (e.g., we eat together at dinner time, we go to bed at night). Only older children appeared to draw on a more adult-like understanding of life as some kind of "vital force," grounded in and governing bodily processes and extending to a variety of entities that are quite different from humans. Over the course of early and middle childhood, Carey argued, children exchange an "animistic" theory of life for another, "vitalistic" one—in the process, demarcating the conceptual domain of living bodies ("lay biology") as separate from the conceptual domain of human social interactions ("lay psychology").

Meanwhile, under the unifying label of "theory of mind," thousands of studies have documented major improvements over the course of early and middle childhood in children's abilities to take others' perspectives, predict and explain people's emotions, represent false beliefs, and integrate representations of beliefs and intentions in evaluating moral responsibility (for reviews, see Flavell, 1999; Wellman, 2015). These studies offer overwhelming evidence that becoming a sophisticated reasoner—and particularly a sophisticated social reasoner—requires substantial refinement of one's representations of others' experiences, abilities, beliefs, desires, and so forth.

Studies in these traditions have generally focused on changes in children's beliefs about the world (e.g., Does the child believe that the moon has intentions?); or on changes in children's intuitive theories (e.g., Does the child draw on social-psychological or biological causes when explaining an entity's behavior? Do they include beliefs, in addition to desires, in their representations of the mental states that drive people's actions?). For the most part, however, these studies do not address the ontological structural underlying children's beliefs, explanations,

and behaviors. Terms like "intention," "belief," and "desire" abound in our scientific theories of conceptual development—but how well do these terms correspond to children's own understanding of different aspects of mental life (or, for that matter, adults' understanding)? Although the field of developmental psychology has made great progress investigating changes in children's beliefs and intuitive theories about the mind, we have very little understanding of changes in *conceptual structure* in this domain.

Key insights from social psychology. Meanwhile, in social psychology, a brief but highly impactful paper published in *Science* in 2007 sparked a surge of interest in adults' representations of mental life, in a field that has come to be known as "mind perception." H. Gray, K. Gray, and Wegner (2007) presented the results of a large-scale study in which adult participants were asked to compare the mental capacities of a variety of target characters (e.g., adults, children, a dead person, a chimpanzee, a frog, a robot, God). From these assessments, the authors derived two axes of variability in participants' responses: a dimension they called "experience," which captured the extent to which participants considered a character to be capable of hunger, fear, pain, pleasure, rage, desire, personality, consciousness, pride, embarrassment, and joy; and a dimension they called "agency," which captured the extent to which participants considered a character to be capable of self-control, morality, memory, emotion recognition, planning, communication, and thought. They termed these the "dimensions of mind perception."

In the dissertation (Chapter I), I provide a detailed critique Gray et al.'s methods and interpretation in the spirit of seeking to deepen the field's understanding of their findings—and as a foundation for my own work, which was designed and executed with these considerations in mind. Here, however, I highlight three contributions from this work that I adopted and expanded upon in the current studies.

First, *a description of "mind" as a multidimensional construct*. Perhaps the most radical—and under-valued—contribution of Gray et al.'s work on mind perception was the very premise of their study: that people's understanding of mental life might be structured along multiple, meaningful "dimensions." Gray et al.'s study identified a level of conceptual organization intermediate between the broad concept of "mind" and the narrow definitions of individual mental states (e.g., "joy," "vision," "belief," "desire"). In so doing, it introduced the idea—absent from the majority of relevant work in cognitive and developmental psychology—that such a conceptual organization might exist, that it can be identified empirically, and that it is worth studying.

Second, *an expansive view of mental life*. Gray et al. included an unprecedentedly wide variety of "mental capacities" in their study—not only the cognitive and emotional abilities that are commonly targeted in studies of reasoning about mental states, but also less common examples of mental states, such as physiological sensations and capacities for emotion recognition and moral reasoning. I argue that this expansive scope is critical for getting a holistic sense of this complicated domain.

Third—and perhaps most critically—*a bottom-up approach*. Gray et al. did not generate a theory of "dimensions of mind perception" a priori and set out to confirm or disconfirm this theory; neither did they ask participants to articulate their own conceptualization of this domain. Instead, they relied on an unsupervised learning algorithm to reconstruct participants' representations of a conceptual domain "from the bottom up."

Such bottom-up analyses are rarely portrayed by social psychologists as bearing on "cognitive" topics like conceptual representations. But I argue that unsupervised learning techniques and other bottom-up statistical analyses have tremendous potential to advance our understanding of conceptual representations of mental life and their development over childhood. First, bottom-up approaches have the potential to elucidate the kinds of deep conceptual structures that are difficult for participants to report on directly. Second, the conceptual structures revealed by bottom-up approaches can, in principle, differ dramatically from the assumptions of a research team. These are especially compelling features in the domain of conceptual change—in which, by definition, participants struggle to introspect and articulate their reasoning, and findings often conflict with adult intuitions (including the intuitions of researchers). In contrast to a top-down approach, in which the researcher must anticipate in advance one or more alternatives forms that this representation must take— and, if they are interested in its development, one or more dimensions along which this representation might vary over childhood—a bottom-up approach allows conceptual structures, and changes in these structures, to emerge organically, and to vary along both anticipated and unanticipated axes of comparison.

Three Key Questions, Inspired By the Ancients

Human inquiry into the nature and structure of the mind, soul, or psyche is of course much older than the fields of cognitive, developmental, and social psychology that provide the theoretical underpinnings of this dissertation. Such topics were addressed by many of the philosophers and spiritual leaders of antiquity. I am particularly indebted to these ancient philosophers for a certain ontological bent toward this topic—which differs from mainstream approaches to conceptual representations in the modern empirical study of cognitive development.

In this dissertation, I address three key ontological questions, beginning in each case with adults and then turning to development in early and middle childhood:

1) What are the conceptual units that anchor representations of mental life? (Chapter III)
2) What is the structural organization governing the relationships among these conceptual units? (Chapter IV)
3) How do people deploy their conceptual representations of mental life to reason about specific entities in the world—namely, animate beings vs. inanimate objects? (Chapter V)

This three-pronged ontological approach parallels key aspects of ancient philosophical inquiries into the nature of the mind. In *De Anima*, for example, Aristotle describes a theory of the soul that addresses (1) the distinct faculties of the soul (nutrition, perception, reason, and perhaps desire); (2) the hierarchical relationships among these faculties (the presence of reason in an entity implies the presence of perception and nutrition, but the reverse is not true; likewise, the presence of perception implies the presence of nutrition, but the reverse is not true); and (3) the way this structure makes sense of the variety of beings in the world (plants have only a nutritive soul; non-animals have both nutritive and perceptual souls; and only humans have nutritive, perceptual, and intellectual souls). (See Chapter I for parallel examples from Plato and the Buddha.)

The content of these ancient theories and the modes of inquiry these thinkers employed differ dramatically from what we would recognize as theories of "the mind" in twenty-first century psychological science—but I have found their shared, three-pronged ontological approach to be a useful way to structure my exploration of ordinary people's understanding of the mind.

Empirical Approach

I examine conceptual representations of mental life by documenting participants' mental capacity attributions to a wide variety of familiar entities that might be perceived to vary in their mental lives, including humans, non-human animals, technologies, and inert objects. The current studies were designed to capture participants' beliefs about the co-occurrence of a diverse range of mental capacities: When someone indicates that some entity has one capacity (e.g., for pain, happiness, or memory), what other capacities does that person tend to attribute to that entity? The goal of these studies was to facilitate participants' engagement with deep questions about the nature of mental life—in particular, the similarities, differences, and relationships among different mental capacities—through simple questions grounded in concrete, real-world examples.

In each study, every participant assessed 1-2 target characters (e.g., a beetle, a robot, a bird, a goat, a computer, a human, 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 a random order).

I use the variability in these mental capacity attributions to shed light on the three aspects of conceptual representation that are the focus of the current research.

First, I use the *covariance* of mental capacity attributions to identify the *conceptual units* among participants of different ages via exploratory factor analysis (Chapter III). For example, if participants who endorsed happiness also tended to endorse capacities for sadness and embarrassment (regardless of which target character they assessed), I take this as evidence that happiness, sadness, and embarrassment constitute a suite of mental capacities that are closely associated with the same underlying "conceptual unit."

Second, I assess *asymmetries* in mental capacity attributions in order to reveal the *structural organization* of these units (Chapter IV). For example, if many participants endorsed capacities associated with Conceptual Unit A without endorsing capacities associated with Conceptual Unit B, but very few participants did the reverse, I take this as evidence that Conceptual Unit A might be considered more basic or fundamental than Conceptual Unit B, or a prerequisite for Conceptual Unit B.

Third, I examine which mental capacities were attributed to which target characters in order to characterize the *deployment* of these conceptual representations in reasoning about entities in the real world (Chapter V). For example, if participants who assessed the mental capacities of birds, goats, and humans shared one general pattern of mental capacity attributions, while participants who assessed robots and computers shared another pattern, I take this as evidence that conceptual representations of mental life play a role in structuring representations of and interactions with different classes of beings in the world (namely, animates vs. inanimates).

Summary of Key Findings

The seven studies of US adults included in this dissertation (total N=1862) converge on a clear picture of US adults' representations of mental life. These representations are anchored by a distinction between three fundamental conceptual units: BODY (characterized by physiological sensations related to biological needs, such as hunger and pain), HEART (characterized by social-emotional abilities like happiness and guilt), and MIND (characterized by perceptual-cognitive abilities to take in and store information about the environment, such as vision and memory).[1] In terms of structural organization, BODY and MIND occupy a more basic or fundamental position in this conceptual structure, together supporting or otherwise giving rise to HEART. Finally, in deploying this folk ontology to reasoning about beings in the world, US adults distinguish animate beings from inanimate objects not only in that their mental capacities are, on average, superior (especially in the BODY domain)—but also in that their mental capacities are more variable across specific entities and more correlated across domains (BODY, HEART, and MIND). Meanwhile, three studies of 4- to 9-year-old US children (total N=493) suggest that, over the course of early and middle childhood, US children's representations of mental life undergo substantial development in all three of these respects.

Together, these have given rise to the following theory of conceptual development in this domain; see Figure 1 for a visual depiction.

First, in terms of *conceptual units*, young children have access to a more limited set of conceptual units than US adults; by the preschool years (roughly 4-5 years of age), US children make a broad distinction between the more visceral sensations of the BODY and the more cognitive abilities of the MIND, but have no notion of social-emotional abilities as a third, unified class of mental states. Over the course of early childhood, the set of conceptual units available to US children expands in number as HEART emerges as a distinct construct; each of these conceptual units also undergoes further refinements in its content and size. The set of conceptual units reaches an adult-like state sometime in the elementary school years, early enough to appear "mature" in a snapshot of 7- to 9-year-old children.[2]

Even by the preschool years, however—well before these conceptual units are fully mature—US children demonstrate a nascent awareness of the *organizational structure* that characterizes US adults' representations: By roughly 4-5 years of age, children already consider physiological sensations (BODY) to be particularly basic or fundamental aspects of mental life, and they quickly come to see perceptual-cognitive abilities (MIND) as roughly equally "basic." The social-emotional abilities of the HEART are already perceived to be *less* basic, i.e., to occupy a different position in the hierarchical structure that characterizes this conceptual domain (even though they are not yet perceived as constituting a unified third construct distinct from BODY and MIND). These structural relationships become increasingly strict with age: Eventually, this hierarchical structure admits of virtually no exceptions—it governs mental capacity attributions

[1] These findings were first presented in the *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (Weisman et al., 2016) and subsequently published in *PNAS* (Weisman et al., 2017a).

[2] These findings were first presented in the *Proceedings of the 39th and 40th Annual Meetings of the Cognitive Science Society* (Weisman et al., 2018, 2017b).

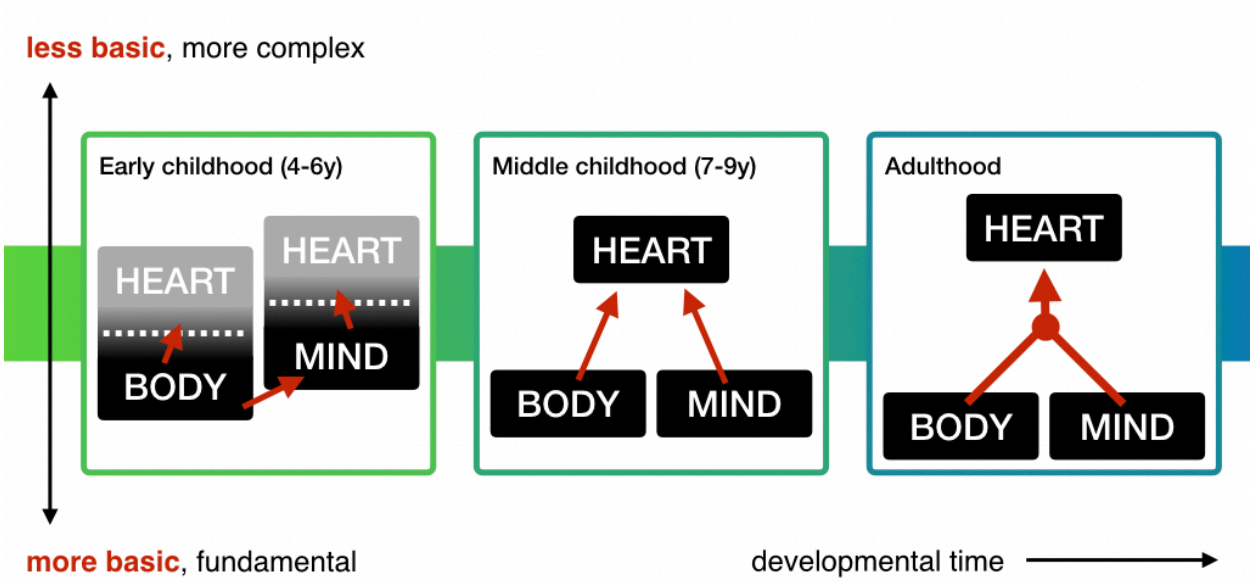

Figure 1: A visual depiction of my working theory of conceptual development in representations of mental life in the modern US context, featuring snapshots at three points in development. Conceptual units are depicted as black boxes. Structural relationships are depicted as red arrows, with the arrowhead pointing to the less "basic" unit. The red node that these arrows pass through in the "Adulthood" snapshot represents the perceived joint dependency of HEART on BODY and MIND.

to all kinds of target entities, for all individual observers—but even by middle childhood US children have not yet come to this degree of strictness. The final element of this relational structure to emerge is the idea that HEART is not only dependent on both BODY and MIND, but *jointly* dependent on their combination. This understanding of joint dependency may emerge from one or more intuitive theories, such as a theory of how emotions work (e.g., that affective experiences have both physiological and cognitive components).

Finally, in terms of the deployment of this concept—from early in childhood, US children's distinction between animate beings vs. inanimate objects includes an understanding, shared with US adults, that animate beings are generally more likely to have any kind of mental capacities than inanimate objects. The BODY, in particular, is the primary axis of this distinction throughout childhood and in adulthood, with this distinction increasing in size and reliability over early and middle childhood (and perhaps beyond). Eventually, US children come to believe that only animates, but not all animates, have social-emotional abilities—i.e., that biological animacy is necessary, but not sufficient, for HEART. Likewise, children eventually come to believe that most animates, but also some inanimates, have perceptual-cognitive abilities—i.e., that biological animacy is broadly sufficient, but not necessary, for MIND. These adjustments to the conceptual connections between mental life and animacy result in general decreases in attributions of HEART, and increases in attributions of MIND. Eventually, children come to share US adults' intuitions (which might be considered "over-hypotheses"; Goodman, 1955) that animate beings are distinct from inanimate objects not only in that their mental capacities are, on average, superior (especially in the BODY domain)—but also in that their mental capacities are more variable across specific entities, and more correlated across domains (BODY, HEART, and MIND).

Implications for Social Reasoning

In describing BODY, HEART, and MIND as "conceptual units" I have drawn primarily on the language of ontology, presenting these three aspects of mental life as component parts of a larger concept, or as categories of mental capacities. In the final pages of the dissertation, I propose a different way of thinking about BODY, HEART, and MIND: as distinct *modes of social reasoning*. I draw on the metaphor of BODY, HEART, and MIND as offering different "lenses" through which an observer might "view" another being's behavior—each lens being associated with a different set of knowledge, concepts, and theories about one aspect of a being's existence and identity, thereby making available to the observer specific ways of interacting with that being.

Through the lens of the BODY, an observer understands a being as an animal, subject to biological needs and motivation to act to satisfy these needs (e.g., Carey, 1985; Wellman & Gelman, 1992). Through the lens of the HEART, an observer understands a being as a social partner, with social relationships, group affiliations, and moral status (e.g., Hamlin et al., 2013; Spelke et al., 2013). And through the lens of the MIND, an observer understands a being as an intentional agent, who takes in, stores, and makes use of information about its surroundings in order to achieve certain goals or end-states; this lens is most consistent with traditional views of representational "theory of mind" (e.g., Wellman & Gelman, 1992; Wellman & Woolley, 1990).

In any given interaction an observer might draw on one or more of these lenses to predict and explain another being's behaviors. Indeed, another gloss on this proposal would be that BODY, HEART, and MIND pick out *three parallel lay theories* of motivated action, each akin to the "BELIEF + DESIRE = ACTION" framework fundamental to many psychological theories (Dweck, 2017; Gross, 2015; Schank & Abelson, 1975; Wellman & Woolley, 1990). Each lens, however, offers a different gloss on this basic calculation, thereby facilitating a different mode of interaction between the observer and the observed.

From this perspective, the empirical results described in this dissertation raise the possibility that the "animal" lens associated with BODY and the "agent" lens associated with MIND might be more fundamental, mastered earlier in development, and perhaps more common across cultural settings[3] than the "social partner" lens associated with HEART—which might be differentiated and refined over an extended period of personal experience and enculturation. In other words, over development children may not only acquire more sophisticated ways of understanding mental life, but additional ways of understanding and interacting with others.

[3] Work building directly on the current studies suggests that the same pattern of similarities and differences that characterized the development of "conceptual units" in the US (Chapter III) also characterizes comparisons across diverse cultural settings in the US, Ghana, Thailand, China, and Vanuatu, as well as development within each of these settings: A distinction between BODY vs. MIND is widely shared, but there are many different ways of understanding how emotions fit in (Weisman et al., ms. under review).

References

Carey, S. (1985). *Conceptual Change in Childhood*. MIT Press.

Dweck, C. S. (2017). From needs to goals and representations: Foundations for a unified theory of motivation, personality, and development. *Psychological Review*, *124*(6), 689–719.

Flavell, J. H. (1999). Cognitive development: Children's knowledge about the mind. *Annual Review of Psychology*, *50*, 21–45.

Gelman, R., & Spelke, E. S. (1981). The development of thoughts about animate and inanimate objects: Implications for research on social cognition. In J. H. Flavell & L. Ross (Eds.), *Social Cognitive Development: Frontiers and Possible Futures* (pp. 43–66). Cambridge University Press.

Goodman, N. (1955). *Fact, Fiction, & Forecast*. Harvard University Press.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619.

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, *26*(1), 1–26.

Hamlin, J. K., Ullman, T., Tenenbaum, J., Goodman, N., & Baker, C. (2013). The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model. *Developmental Science*, *16*(2), 209–226.

Piaget, J. (1929). *The Child's Conception of the World*. Routledge & Kegan Paul Ltd.

Schank, R. C., & Abelson, R. P. (1975). Scripts, plans, and knowledge. *Proceedings of the 4th International Joint Conference on Artificial Intelligence*, 151–157.

Spelke, E. S., Bernier, E. P., & Skerry, A. E. (2013). Core social cognition. In M. R. Banaji & S. A. Gelman (Eds.), *Navigating the Social World: What Infants, Children, and Other Species Can Teach Us* (pp. 11–16). Oxford University Press.

Weisman, K., Dweck, C. S., & Markman, E. M. (2017a). Rethinking people's conceptions of mental life. *Proceedings of the National Academy of Sciences*, *114*(43), 11374–11379.

Weisman, K., Dweck, C. S., & Markman, E. M. (2018). Folk philosophy of mind: Changes in conceptual structure between 4-9y of age. *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*.

Weisman, K., Dweck, C. S., & Markman, E. M. (2017b). Children's intuitions about the structure of mental life. *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, 2017.

Weisman, K., Dweck, C. S., & Markman, E. M. (2016). Varieties of experience: A new look at folk philosophy of mind. *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*, 2741–2746.

Weisman, K., Legare, C.H., Smith, R.E., Dzokoto, V.A., Aulino, F., Ng, E., Dulin, J.D., Ross-Zehnder, N., Brahinsky, J.D., & Luhrmann, T.M. (ms. under review). Concepts of mental life among adults and children in five cultures.

Wellman, H. M. (2015). *Making Minds: How Theory of Mind Develops*. Oxford University Press.

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

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, *35*(3), 245–275.