A Unified Account of General Learning Mechanisms and Theory–of–Mind Development

THEODORE BACH

Abstract: Modularity theorists have challenged that there are, or could be, general learning mechanisms that explain theory-of-mind development. In response, supporters of the 'scientific theory-theory' account of theory-of-mind development have appealed to children's use of auxiliary hypotheses and probabilistic causal modeling. This article argues that these general learning mechanisms are not sufficient to meet the modularist's challenge. The article then explores an alternative domain-general learning mechanism by proposing that children grasp the concept *belief* through the progressive alignment of relational structure that occurs as a result of structural-comparison. The article also explores the implications of the proposed account for Fodor's puzzle of conceptual learning.

1. Introduction

Supporters of the 'scientific theory-theory' (STT) account of theory-of-mind (ToM) development claim that children employ domain-general learning mechanisms in order to construct cognitive-behavioral theories in ways analogous to scientific theory construction (Bartsch and Wellman, 1995; Gopnik and Meltzoff, 1997; Gopnik and Schulz, 2004). An important objection to this view, often raised by supporters of modular accounts of ToM development, is that cognitive science has yet to disclose domain-general learning mechanisms that can explain ToM development. Fodor even argues for this point on conceptual grounds, claiming that 'there literally isn't such a thing as the notion of learning a conceptual system richer than the one that one already has' (Fodor, 1980, p. 148). Advocates of STT such as Gopnik and Schulz are also aware of the need to identify learning mechanisms, stating that 'it is all very well to suggest that children's learning mechanisms are analogous to scientific theory-formation. However, what we would really like is a more precise specification of the mechanisms that underlie learning in both scientists and children' (Gopnik and Schulz, 2004, p. 371).

In addressing this concern, it is somewhat surprising that supporters of STT have limited their attention to the mechanisms of probabilistic causal modeling and the incorporation of auxiliary hypotheses. In Section 2 I critically evaluate

Address for correspondence: Department of Philosophy, Bowling Green State University Firelands College, 1 University Drive, Huron OH. 44839, USA.
Email: theodorebach@gmail.com

these two STT proposals for how general learning mechanisms explain transitions between stages of ToM. Sections 3 and 4 propose an alternative general learning mechanism—progressive relational alignment through structural-comparison—that avoids the problems discussed in Section 2. In Section 3 I describe the learning mechanism of progressive alignment through structural-comparison, and I discuss how this mechanism converts implicit relational knowledge into explicit relational knowledge. Section 4 describes how children develop the concept *belief* by employing structural-comparison in order to make explicit the cognitive-behavioral relational system that defines *belief*. Section 5 considers the possibility that young infants have a precocious and tacit grasp of the concept *belief*, and Section 6 concludes.

2. Scientific Theory–Theory and the Development of Belief–Desire Psychology

According to STT, children rationally progress through stages of ToM development on route to their grasp of belief. Bartsch and Wellman (1995) outline three main stages of theoretical skill during this period.[1] Children are *desire-psychologists* from about the age of 2, at which time they: understand simple emotional and perceptual states; have knowledge of desire; understand desire as a subjective and internal state (but not in representational terms); and understand that people 'act so as to satisfy their desires'. When children become *desire-belief psychologists* (around the age of 3) they grasp that cognitive states have internal, representational content. That is, they learn that objects, events, state-of-affairs, etc., are represented in the minds of individuals. Desire-belief psychologists have some grasp that representations can be false, but they do not yet refer to these representations in the explanation and prediction of behavior. Instead, they are restricted to the desire-action schema. At around the age of 4 children graduate to a *belief-desire psychology*. Here they replace the desire-action schema with the belief-desire action schema which encodes that 'persons act *on the basis of their beliefs* to achieve their desires' (Bartsch and Wellman, 1995, p. 154). At this point they regularly pass the (direct) false–belief task and have grasped the concept *belief*.

Bartsch and Wellman explain how children progress through these stages in terms of their ability to incorporate 'auxiliary hypotheses' about cognitive-behavioral relationships. When desire-psychologists confront phenomena not readily explicable

---

[1] Bartsch and Wellman base their developmental timetable on direct false-beliefs tasks (see also Wellman, Cross and Watson, 2001). Studies that employ spontaneous-response tasks (see Baillargeon, Scott and He, 2010) suggest an earlier timetable. The ensuing discussion will assume that data from direct false-belief tasks accurately track the onset of belief-desire psychology. Section 5 examines the possibility that young infants first understand belief-desire psychology tacitly.

by their non-representational concept they will develop auxiliary hypotheses about representational states. Similarly, when desire-belief psychologists confront phenomena not readily explicable by the desire-action schema they will develop auxiliary hypotheses about the explanatory role of belief. These explanatory posits remain peripheral to the general theory of mind and behavior, but eventually they are given central status:

> In the process of utilizing desire-psychology to predict and explain people's actions and emotional reactions, children may encounter certain puzzles or explanatory failures. For example, the child may notice that he and another person have similar desires but engage in two different actions … At first, an appeal to representational states is recruited only when desire psychology breaks down or in order to provide a mental state explanation that is consistent and relevant. Consideration of what the actor wants still drives children's explanatory efforts … Consideration of beliefs as framing the actor's desires, not merely existing as auxiliary to desire psychology, is the key notion behind belief-desire reasoning … What began as an auxiliary hypothesis becomes increasingly typical and finally prototypical of all mental states (Bartsch and Wellman, 1995, pp. 171−2).

This account may be accurate, but it is lacking in an important respect. Specifically, it does not tell us where the auxiliary hypotheses come from. It tells us in what context such hypotheses are needed, but it does not say how they are developed. Here I think Fodor's question about concept acquisition is apt: from what resources does the desire-psychologist draw in order to construct an auxiliary hypothesis about causally efficacious belief states? If there are such resources, then the resulting view threatens to collapse into nativism about the concept *belief* .

To illustrate this collapse, consider that several proponents of modular explanations of ToM development (Scholl and Leslie, 1999; Leslie, Friedman and German, 2004) claim that 'higher-order' cognitive processes, for example theorizing, facilitate ToM development. However, they maintain that the input for these processes—the metarepresentational concepts *belief* and *desire*—have an innate basis in the ToM module. Without a domain-general account of hypothesis discovery, Bartsch and Wellman's account of ToM development is consistent with (if not better explained by) this modular account of innate metarepresentational concepts. In order to distance themselves from nativism, then, Bartsch and Wellman must do more than describe the contexts in which mentalistic hypotheses are explanatorily useful and how such hypotheses are integrated into more general cognitive-behavioral theories. Bartsch and Wellman must also show, *in reference to domain-general learning mechanisms*, how mentalistic hypotheses are developed in the first place.

The recent appeal by supporters of STT to the general learning mechanisms of probabilistic causal modeling also fails to address the nativist's challenge. Gopnik explains children's theory construction in reference to children's ability to reason probabilistically and construct causal models. She expresses the basic idea this way:

> The new idea is to formally integrate structured hypotheses, such as gram-
> mars, hierarchies or causal networks, with probabilistic learning techniques,
> such as Bayesian inference. The picture is that children implicitly consider many
> hypotheses and gradually update and revise the probability of those hypotheses
> in the light of new evidence (Gopnik, 2011, p. 129).

Bayesian learning is a mechanism through which children can weigh the explana-
tory merits of one theory (e.g. desire–belief psychology) against another theory (e.g.
belief–desire psychology) as they acquire new evidence through observation and
intervention. Through this learning technique children are able to construct more
accurate and predictive causal maps of the cognitive-behavioral domain. But as sev-
eral critics have pointed out, such accounts fail to explain how learners acquire the
initial variables that serve as input to these learning techniques. In other words, the
accounts do not explain how children discover hypotheses (Boroditsky and Ramscar,
2001; Goldman, 2006; Christie and Gentner, 2010). This is especially meaningful
with respect to the ToM debate, because explanation of the development of men-
talistic hypotheses is precisely what distinguishes STT and modular accounts.

In order to address concerns about hypothesis discovery, supporters of Bayesian
explanations of learning have recently made a distinction between 'latent' and
'explicit' hypothesis spaces. According to Perfors, Tenenbaum, Griffiths and Xu
(2011):

> … if people have the capacity to represent some given hypothesis, we say it can
> be found in their *latent hypothesis space* … we contrast this with hypotheses that
> may be explicitly considered or evaluated—the hypotheses that can be actively
> represented and manipulated by the conceptual system—which we refer to as
> the *explicit hypothesis space* (Perfors *et al.*, 2011, p. 315).

They demonstrate the distinction by analogy to an English typewriter. Each doc-
ument that the typewriter is *logically capable* of producing—'which includes things
like *The Tempest* and does not include, say, a Vermeer painting or a poem written
in Russian' (p. 315)—is one of the infinite hypotheses found in the latent hypoth-
esis space. 'The documents that have actually been typed out so far make up the
explicit hypothesis space' (p. 315). Their suggestion is that learning consists in the
entire process of 'moving' hypotheses from the latent space into the explicit hypoth-
esis space where they can be evaluated. Latent hypotheses are innate, but because
they are logical capacities rather than structured representations, the concession to
nativism is slight and unproblematic.

There are two problems for this proposal that are worth examining. First, a tem-
poral ambiguity affects how we should understand the 'capacity' of a system that
changes as a result of learning and development. For example, we might deny that
newborn infants possess latent hypotheses about weak nuclear forces and prebiotic
evolution insofar as their brains and conceptual structures do not currently have
the capacity to token such hypotheses. On the other hand, these very minds may

eventually entertain such hypotheses, in which case it would appear to follow that the neonate does possess latent hypotheses about nuclear forces and primordial conditions. The analogy to the typewriter, which is a non-developing system, breaks down at precisely this point.

Second, and related, Perfors *et al.*'s proposed distinction between hypothesis spaces faces a dilemma. Either latent hypotheses refer to selectable and manipulable tokens, or they refer to the logical capacities of a system. If they refer to selectable and manipulable tokens, then learners possess extensive and innate propositional knowledge, in which case the concession to nativism is total rather than slight. If they refer to the logical capacities of a system, then we are back at the original and problematic question, namely, how a learning system is able to actualize its developmental capacities in order to make available knowledge structures that can then be targets for explicit (e.g. Bayesian) reasoning. The worry is that the authors emphasize capacity in order to distance themselves from radical nativism, but then emphasize manipulable tokens in order to explain the conversion of latent to explicit hypotheses. Consider the author's claim that 'learning … would never involve new hypotheses being added to the latent hypothesis space' (p. 315). If the term 'hypotheses' in this passage refers to *capacities*, then this claim is uncontroversial. At the same time, this interpretation immediately shifts the explanatory burden to provide an account of how latent 'hypotheses', *qua* capacities, are converted into explicit hypotheses—how capacities defined by logical possibility become, e.g. symbolic knowledge structures. When the authors sketch such an account, they shift from talking about latent hypotheses in terms of capacities to talking about them as manipulable, selectable tokens: 'Hypothesis generation would describe the process by which hypotheses move from the latent space to the explicit space—the process by which our typist decides what documents to produce' (p. 315). However, it would appear to be a category mistake (Ryle, 1949) to say, even metaphorically, that *capacities* 'move' between cognitive spaces or are 'selected' for production.

What moral should we draw from this discussion? I suggest the following: supporters of STT must identify a general learning mechanism for hypothesis discovery that can supplement the processes of probabilistic causal modeling and/or auxiliary hypothesis assimilation. Without such a unified account of domain-general learning mechanisms, defenders of STT remain vulnerable to nativist critiques. In subsequent Sections I argue that relational abstraction through structural-comparison and progressive alignment is such a mechanism. Along the way I describe how this view strikes a middle ground with respect to Fodor's puzzle about conceptual learning.

3. Structural–Comparison and the Relational Shift

In order to grasp fully the belief-concept, children must conceptualize what philosophers term the belief-desire law (or belief-desire schema): *Persons who desire the content of a representation P and believe that action A will bring about the realization of P are*

disposed to perform action A. In order to determine how and when children obtain this knowledge structure, it is worthwhile to ask the following two questions:

(a) What *type* of knowledge structure is the belief-desire schema?
(b) Are there any well-understood learning processes through which children acquire knowledge structures of this type?

In providing an answer to question (a), it is often overlooked that the belief-desire law is a relational category. The members of relational categories are united on the basis of shared relational features, e.g. causal or functional properties (Markman and Stillwell, 2001; Gentner, 2005; Gentner and Kurtz, 2005), and they are multiply realized by objects. In contrast, object categories (dog, grass, rock) are marked by common intrinsic and/or perceptual features and are not multiply realizable. We can further distinguish between relational system categories (robbery ecosystem, democracy) and relational-role categories (thief, predator, president).

Question (b) can now be rephrased as follows: are there any well-understood processes through which children develop relational category knowledge? A central appeal of Gentner and colleagues' *structure-mapping theory* is its explanation for how learners develop relational category knowledge.[2] The core idea is that comparison, which involves the structural-alignment of two representations, induces an epistemic focus on relations while demoting the epistemic importance of object-realizers. Such highlighted relations can then be abstracted and serve as the basis for relational concepts that describe relational categories. Below I explain and provide examples of this important developmental process. But before doing so, it will be important to explore two points. First, I explain the unique developmental challenge that relational categories present to learners. Second, and in order to forestall confusion, I will explain an important respect in which analogical learning traditionally construed does *not* bear centrally on the development of relational category knowledge.

Prior research indicates a learning bias during development against relational categories and in favor of object categories. According to the MacArthur Communicative Developmental Inventory, of the 296 nouns that 8- to 16-month-olds understand, 93 per cent are object nouns and 7 per cent are mixed object-relational nouns (Gentner, 2005, p. 251). When children do learn relational terms, they are initially treated as object terms (*ibid*. p. 251). An important difference between object and relational categories explains this bias. Object categories have rich overlap of superficial features that are readily detectable by children. In contrast, the axes of similarity that bind relational category members are less obvious and detecting them

---

[2] Given a high degree of interdisciplinary convergence on foundational issues and models, the study of analogical and similarity cognition is widely viewed as one of the 'success stories' of cognitive science (Forbus *et al.* 1998, p. 231). This essay provides an application of Gentner and colleagues' structure-mapping theory—the most widely accepted model of analogical cognition. See Bach, 2012 and Gentner and Colhoun, 2010, for general overviews of structure-mapping theory.

requires conceptual sophistication. As Gentner puts it, object categories are 'out there', natural partitions in the world. The extensions of relational categories, on the other hand, are diffuse and varied. For example, it is difficult to abstract the concept *gift* because there are no intrinsic or perceptual features that all gifts must exemplify. Eventually, children make a 'relational shift' and conceptualize a domain in terms of relational categories rather than, or in addition to, object categories. For example, they grasp that uncles must exemplify the relational property 'male sibling of a parent' rather than the attribute 'nice man with a pipe' (Gentner and Kurtz, 2005), or that any object can be a gift if it plays the correct role in a relational schema of gift–giving. Structure-mapping theory helps explain how children negotiate this shift and overcome their initial bias towards objects and attributes.

Now, a common way to think about analogical learning is in terms of a well understood base representation (e.g. a representation of the solar system), a less well understood target representation (e.g. a representation of the Rutherford atom), and then an inference that is projected from the base to the target (e.g. the inference that electrons revolve around the nucleus of the Rutherford atom). Such projective analogies are of course important for learning. However, because they require antecedently understood base representations, projective analogies generally presume rather than explain the development of relational knowledge. Fortunately, there are other and more dynamic aspects of analogical cognition that better explain the discovery of relational category knowledge and which offer better resources for addressing the nativist challenge. Below I describe these processes, which include relational focus, progressive alignment, and relational schema abstraction.

There is now considerable evidence that the general activity of comparison, which involves the juxtaposition and alignment of two representations in working memory, promotes attention towards relational features while epistemically 'demoting' the importance of object-level features (attributes).[3] When learners consider one exemplar they are likely to focus on non–relational features (for example the appearance of the object). But when there are two relationally similar exemplars and a subject compares them, the activity of alignment will highlight common relational structure rather than shared attributes.[4] Metaphorically, aligning two representations

---

[3] Following Gentner and colleagues, the surface elements of a representation include 'entities' and 'attributes'. Entities reference individuals and particular objects, and attributes are predicates that take one argument, e.g. RED (ball). The structural elements of a representation are relations, which are predicates that take at least two arguments. Relations subdivide between first-order relations and higher-order relations. First-order relations take objects as arguments, e.g. COL-LIDE (car, wall). Higher-order relations take propositions as arguments, for example, 'The car colliding with the wall caused the bricks to strike the windshield', or, CAUSE [COLLIDE (car, wall), STRIKE (bricks, windshield)].

[4] Not only are relations preferred over attributes, but systematic relations—those that participate in a wider system of relations—are preferred over isolated relations. This 'systematicity constraint' addresses Goodmanian concerns about proliferating relational similarities. For example, the systematicity principle accurately predicts that aligning representations for the Rutherford atom and the solar system will highlight the common relation *The planets/electrons REVOLVE*

causes a common relational structure to 'pop-out'. Gentner *et al.* (1997) term this the 'comparison as X-ray phenomena'. Highlighted relations can then be abstracted and serve as the basis for relational concepts that describe relational categories.

There are several lines of empirical support for comparison-based relational focus and abstraction. Using Duncker's classic radiation problem, Gick and Holyoak (1980, 1983) observed a relationship between comparison and relational transfer. The radiation problem asks subjects to devise a method for using rays to destroy a patient's tumor while not also destroying the healthy tissue that surrounds the tumor.[5] The solution to the problem is to arrange for several low intensity rays to converge simultaneously and from different directions on the tumor: the individual rays have insufficient intensity to destroy the healthy tissue but their combined effect is sufficient to destroy the tumor. Subjects discover this solution on their own approximately 10% of the time. If given a prior, relationally similar story (about a general planning to capture a centrally-located fort that is protected by mines over which small but not large groups of soldiers can safely travel), *and also given the convergence solution to that story problem*, subjects go on to solve the radiation problem about 30% of the time. The marginality of this increase is surprising. However, Gick and Holyoak found that when given a hint to *use the prior story*, subjects' success improves to about 85%. Here, the direction to use the prior story invited a comparison between the two relationally similar problems. This comparison induced a relational focus, facilitating the abstraction and transfer of a relational schema.[6]

One might think that relational focus obtains only in cases like the above—cases for which the compared exemplars share relational features but not attribute features. Importantly, relational focus occurs even when the compared items are superficially (attributionally) similar. Gentner and Namy (1999) provided children with a standard item (a bicycle or a tricycle in one condition) and then asked them to match the standard with either a perceptually similar item (eye-glasses) or a relationally similar item (skateboard). When each standard—the bicycle or the tricycle—was presented

---

AROUND the sun/nucleus. This is because this relation is constrained by the higher order relation *The fact that the X attracts Y CAUSES Y to revolve around X*. On the other hand, because *hotter than* does not participate in an interconnected system of relations, it is not preferred during the mapping process.

[5] Gick and Holyoak gave subjects the following version of Duncker's radiation problem: 'Suppose you are a doctor faced with a patient who has a malignant tumor in his stomach. It is impossible to operate on the patient, but unless the tumor is destroyed the patient will die. There is a kind of ray that can be used to destroy the tumor. If the rays reach the tumor all at once at a sufficiently high intensity, the tumor will be destroyed. Unfortunately, at this intensity the healthy tissue that the rays pass through on the way to the tumor will also be destroyed. At lower intensities the rays are harmless to healthy tissue, but they will not affect the tumor either. What type of procedure might be used to destroy the tumor with the rays, and at the same time avoid destroying the healthy tissue?' (Gick and Holyoak, 1980, pp. 307–8)

[6] See Lowenstein, Thompson and Gentner, 1999, for another example of a study that shows the difference between receiving examples, on the one hand, and comparing examples, on the other.

STANDARD 1

STANDARD 2

PERCEPTUAL MATCH   **CATEGORY MATCH**

Figure 1 *Sample materials from Experiment 2 (from Gentner and Namy, 1999 with authors' permission). For the no-comparison condition children were given just one of the standards. They were told it was a 'blicket' and were asked which of the other items was also a blicket. In the comparison condition children were given two standards and told 'this is a blicket, and this is a blicket, too', and then they were asked which of the other items was also a blicket. Nine other sets of materials (each with two standards, one perceptual match, and one relational match) were also used and achieved similar effects.*

alone, children chose the perceptual match. However, when both standards were presented together and children were invited to compare them, children had a strong preference for the relational category match (see Figure 1). Because the two standards shared the same perceptual similarities with the perceptual match, one might expect stronger preference for the perceptual match when both were presented together. That the opposite occurred is strong evidence for the claim that comparison induced a focus selectively on relational features. Subsequent studies have extended this effect (Namy and Gentner, 2002) as well as controlled for the impact of children's prior knowledge of the relational categories (Christie and Gentner, 2010). The effect

has also been demonstrated in a variety of domains, including learning business negotiation strategies (Lowenstein, Thompson and Gentner, 1999) and the learning of basic spatial relations like symmetry and monotonic change (Kotovsky and Gentner, 1996).[7]

According to structure-mapping theory, as children gain experience in a domain, their increased opportunity for comparison (often guided by language) enables relational abstractions that facilitate their relational shift in that domain. This is an incremental process. Pre-shift, because children are unable conceptually to notice a similarity between two members of a relational-role category that do not share superficial features (e.g. a space-station bridge and a familiar road bridge), they will be unable or unlikely to align their representations for such pairs. This obstructs the type of relational learning just discussed. Fortunately, children can conceptualize as similar, and thus align representations for, relational-role category members that are attributionally similar (e.g. two similar looking road bridges). This perceptual 'hook' allows the importance of relational predicates to slip in through the back door: perceptual similarities trigger the alignment, but the process of alignment privileges the importance of relations that were previously understood only implicitly and contextually. Highlighted relations can then be abstracted and transferred to a more diverse set of object participants. By performing easy and close alignments, then, children can abstract relational knowledge that enables more distant alignments. These more distant alignments can lead to further refinement of relational knowledge, which enables even *more* distant alignments, and so on:

> Consider the example of monotonic change as it might first be learned by a child in a highly concrete context, such as the descending heights of a 'Daddy Mommy Baby' set of dolls. The relational structure of descending size is at first implicit and embedded in the specific family context. At this stage the child would not recognize that the same structure occurs in, say, a set of bowls of decreasing diameter. But if the child is given a close match—say, a different set of descending-size dolls—then the obvious similarities will prompt an alignment process and help to guide it. Miraculously, even such a close alignment can elevate the salience of the common relational structure, thereby potentiating a subsequent more distant match, such as that between the dolls and the bowls. If this process continues—with each new analog clarifying and refining the common structure further—the result can become steadily more abstract (Gentner and Colhoun, 2010, p. 39).

---

[7] An important component of structure-mapping theory is its computational account of relational alignment and relational abstraction. The processing principles of structure mapping, which include alignment, retrieval, and progressive alignment, have each been computationally simulated by the Structure-Mapping-Engine (SME), MAC/FAC, and SEQL respectively. Gentner and Forbus (2011) provide a comprehensive review of these and other computational models of analogical cognition. One noteworthy contribution of these computational models is that they demonstrate how a leaning system that begins 'blind and local'—with no prior programming for the target relational structure—can arrive at the target relational structure.

This process—which Gentner and colleagues term 'progressive alignment'—thus involves a virtuous feedback loop between superficial hooks, comparison-based relational focus, and relational abstraction.

Progressive alignment enables children to abstract relational schemas that are explicit about the essential properties of relational categories. Prior to their relational shift in a domain, children have only an implicit understanding of the relations in that domain. This is because children have conflated relations with specific argument values. After the shift, when children have made explicit the relations that are the essential property of a relational category, they no longer conflate relations with idiosyncratic object-level realizers. For the remainder of this article I will use the following formalization to capture this distinction: *Implicit knowledge* is relational knowledge that is embedded in a particular context or conflated with a particular argument value. In contrast, *explicit knowledge* encodes a relation as multiply realizable by objects. We can also express this difference in representational terms: implicit knowledge involves possession of a relational predicate that will take only one, or a set of superficially similar, argument values, whereas explicit knowledge involves the possession of a relational predicate that can be matched across superficially different argument values.

To demonstrate, a three-year-old who understands that two people can use a telephone to communicate with one another, but who embeds this relational knowledge in a particular exemplar (i.e. a familiar wall phone, or a particular set of adults) has implicit but not explicit relational knowledge of telecommunication. On the other hand, a five-year-old who understands that Bluetooth headsets are used for telecommunication, and who possesses a representational symbol for telecommunication that can take novel arguments, has explicit knowledge of the relation of telecommunication. Progressive alignment and schema abstraction through structural-comparison are thus learning mechanisms that convert implicit knowledge into explicit knowledge. In Section 4 I will examine how this learning mechanism offers distinct responses to nativist challenges about hypothesis search and discovery.

Finally, it is important to be clear on the contexts in which alignment-based learning occurs. Structural-comparisons occur as a result of, or through, domain experience (Gentner *et al.*, 1995), and domain experience has this impact by either culturally or naturally prompting ('inviting') structural-comparison. Cultural invitations occur through common labels or expressions that symbolically juxtapose objects or events. For example, as we saw in the discussion of experiments from (Gentner and Namy, 1999), labeling various devices a 'blicket' across different contexts invites children to perform comparisons that result in the abstraction of a relational category. Importantly, comparing entire sentences has this effect as well (Gentner and Namy, 2004, pp. 557−9; Fisher, 2000). Natural comparison invitations occur through spatio-temporal juxtapositions (compresent or contiguous objects/events), e.g. a child's repetitive act of knocking-over and reassembling a block tower (Gentner, 2005).

4. Using Structural-Comparison to Discover the Belief Concept

Given the substantial literature on ToM development, there has been surprisingly lit-tle discussion of ToM development in terms of alignment-based learning. A notable exception is Baldwin and Saylor (2005). While I agree with Baldwin and Saylor's account on a number of points—particularly the importance ascribed to the role of language in prompting alignment-based learning of mental concepts—the current proposal departs from Baldwin and Saylor's account in several respects. For one, the current proposal models a more expansive role for alignment-based learning in ToM development that incorporates pretense and simulation (Section 4.3) as well as chil-dren's ability to bootstrap their understanding of pictures and *says-that* constructions in order to develop a more generalized conception of *representation* (Section 4.4). Second, the proposal applies structure-mapping principles in order to explain how children discover mentalistic hypothesis*,* and related, how children affect transitions between conceptual stages of ToM development. Indeed, I think that this focus on discovery is the only way to capture the correct role for alignment-based learning in discussions of ToM development. I also contend that it is the best way—perhaps the only way—to provide a unified account of general-learning mechanisms that meets the nativist's challenge. In order to make this application clear, I first remark on Baldwin and Saylor's contrasting application of structure-mapping principles and explain what I view as a mistaken assumption in that application.

An important feature of Baldwin and Saylor's account is their discussion of absent reference. They claim that speaker reference to absent objects, e.g. referencing a non-present dog, will invite children to access their knowledge of previous (successful) dog-references, and that children will then map this knowledge to the current context of absent reference. Given structure-mapping principles, the authors predict that children will thereby focus on the relationship between internal mental states (i.e. *intention*, *attention*) and behavior. Central to this proposal is the claim that:

> The urge to achieve parallel mapping from source to target domain can't be met in any simple way, so an inference may be drawn that enables them to be met at a more abstract level. One such inference that would do the trick is that /dog/ utterances and dog-directed gaze are separate reflections of an unseen, internal dog-related thrust of focus (i.e. intention/attention). Under this inference, /dog/ utterances can reflect dog-related focus even when dogs are absent (Baldwin and Saylor, 2005, p. 134).

As this passage makes clear, it is the discovery of an inference about hidden subjectivity that is doing the heavy lifting of conceptual change on this account. But—and this is the same type of error we encountered in Bartsch and Wellman's appeal to auxiliary hypotheses—this skips over the pressing developmental question of how children came up with this inference in the first place. Reporting that an inference to internal mental states will resolve an explanatory or alignment

problem does little to explain how, or from what premises, such an inference can be drawn.[8] As we saw with Bartsch and Wellman's proposal, two distinct epistemic achievements are conflated: the contexts and learning mechanisms that *expose* a theory's explanatory inadequacy, on the one hand, and the contexts and learning mechanisms through which one *discovers* a solution to the explanatory problem, on the other hand. While it is tempting to build the latter achievement into the former, I maintain that structure-mapping theory motivates only an explanation of the latter. Alignment-based learning occurs gradually and unintentionally through the mundane comparison of high-similarity pairs. This motivates an account for which mentalistic representational resources are discovered 'off stage'—discovered in contexts distinct from those in which they will later be pressed into service. Only after representational resources are made available can they figure in explicit forms of theoretical reasoning (such as Bayesian learning techniques) and provide more comprehensive solutions to the challenges of mentalizing.

The remainder of Section 4 demonstrates this contrasting application of alignment-based learning to ToM development. I first describe the relational-role categories of folk-psychology and then I argue that structural-comparison facilitates the discovery of hypotheses about the relations that define these categories.

4.1 The Relational-Role Categories of Belief–Desire Psychology

I will assume that intentional agents exemplify a functional system that organizes relations between epistemic states (beliefs), goal-directed states (desires), and goal-directed behavior (action). It is because agents exemplify such a system that interpreters can reliably use the belief-desire schema in order to explain and predict intentional behavior. As mentioned in Section 3, relational systems (e.g. democracy, ecosystem) define relational-role categories (e.g. president, predator). Similarly, the belief-desire relational system defines the relational-role categories *desire* and *belief*:

Desire: To desire that P is to have a mental representation that P that plays a certain causal role, namely, that of disposing one to take action A if one were to believe that taking action A would bring it about that P.[9]

---

[8] Analogously, arriving at the realization that one's car has been stolen does little to explain its current location or how one is going to get to work. Or, a fast-ball pitcher's realization that an opposing team cannot hit curve balls does not automatically give that pitcher the ability to throw a curve ball. Perhaps the authors intend to invoke the notion of *re-representation* (Gentner *et al.*, 1995), which occurs when a subject restructures a domain representation in order to improve a relational mapping. If this is correct, more needs to be said about how a domain representation that includes reference to an ontology of internal representations can be restructured from a domain representation that does not include reference to an ontology of internal representations.

[9] This is adapted from Schroeder's discussion of the 'standard view' and 'representational view' of desire—see Schroeder, 2004, p. 17 and p. 24. I also want to be clear that I am not claiming that these definitions exhaust the meaning of the terms 'belief' and 'desire'.

Belief: To believe that 'action A would bring about that P' is to have a mental representation that plays a certain causal role, namely, that of disposing one to take action A if one were to desire that P.

Recalling Bartsch and Wellman's stages of ToM development, children conceptually grasp these relational-role categories only when they are belief-desire psychologists. As discussed in Section 2, current domain-general learning accounts fail to explain children's conceptual discovery of the mentalistic hypotheses required to grasp these categories. Here, I provide concrete demonstrations of how comparison-based learning provides the missing transitional explanation. Specifically, I explain how, assuming Bartsch and Wellman's description of the different stages of ToM development, alignment-based learning enables children to make a *relational shift* in the domain of cognition and behavior and convert implicit cognitive-behavioral relational knowledge into explicit cognitive-behavioral relational knowledge.

I begin with children's transition from desire-belief psychology to belief-desire psychology. Desire-belief psychologists grasp that internally represented desire-contents are causally related to behavior. However, they do not yet possess a knowledge structure that is explicit about the causal relation between *mental representation* and behavior. Instead, they conflate the causal relation between mental representation and behavior with a particular type of mental representation, namely, *desire*. This is analogous to the child who conflates the uncle relation with a particular person or a particular age group. According to the proposed framework for implicit and explicit knowledge, we can say that the desire-belief psychologist's knowledge of the causal relationship between propositional representations and behavior is implicit. The belief-desire psychologist, on the other hand, has made explicit the causal relationship between propositional representation and behavior and can ascribe this relation to objects other than desire. Specifically, the relation can be predicated of *belief*.

Below I model how symbolic juxtapositions of cognitive-behavioral relations invite representational comparisons that facilitate this conceptual transition. In Section 4.3 I model a similar role for spatio-temporal juxtapositions of cognitive-behavioral relations.

4.2 Symbolic Juxtaposition of Cognitive-Behavioral Relations

We can demonstrate the symbolic juxtaposition of cognitive-behavioral relations by examining what Bartsch and Wellman call 'contrastives'. Contrastives are utterances or mini-conversations between a child and an adult during which a mentalistic description is contrasted with reality, fiction, a behavioral outcome, or another mentalistic description. Bartsch and Wellman are interested in contrastives as *indicators* of children's knowledge about the mind. I agree that contrastives indicate knowledge, but I submit that their real value for the study of folk-psychology is their *impact* on development.

Here are some examples of contrastive utterances:

ADULT: Don't.
SARAH (2y 10m):   I want … don't want … I don't want it on [a Band–Aid].
ADULT: Well you gotta have it on. Leave it on.
SARAH: No.

(Bartsch and Wellman, 1995, p. 83)

ADULT: We'll turn it on later.
ABE (2y 10m):   You turn it on later?
ADULT: Yep.
ABE: No. we don't want it on later … I want it on now. You
                                said you will when the cowboys on?
ADULT: I said I'll turn them on later.

(Bartsch and Wellman, 1995, p. 83)

ADAM (3y 9m):   Can you eat snails?
ADULT: Some people eat snails, yes.
ADAM: Why?
ADULT: Because they like them.
ADAM: Mommy, do you want to eat snails?
ADULT: No, we don't think I'd like to eat snails.
ADAM: I don't like to eat snails … People eat snails.

(Bartsch and Wellman, 1995, p. 85)

Contrastives such as the above are symbolic juxtapositions of the causal rela-
tionship between representation and behavior. As such, they should invoke the
comparison of structured representations with desire-action content. Let's take a
closer look at this last contrastive (although any of them would be fine for our
purposes). Two actions are contrasted—eating snails and not eating snails—each
associated with a desire—to eat snails and not eat snails respectively. In order to
apply structure-mapping principles, it is important to keep in mind that relational
abstraction is constrained by the content of the aligned representations, and also that
the content of aligned representations is determined by pre-existing domain knowl-
edge. Assuming the data from direct false-belief tasks (but see Section 5), Adam
(3 years, 9 months) should be in a state of transition from desire-belief psychology
to belief-desire psychology. As discussed above, the desire-belief psychologist's
folk-psychological knowledge is expressed in the representational desire-action
schema—'people act in order to realize the content of their desires'. We can then
describe the compared relational structures as follows:

a. People's desire to eat snails causes them to eat snails.
b. Mommy's desire not to eat snails causes mommy to not eat snails.

Let me remind the reader of one of the basic effects of structure-mapping. When
one exemplar is present, people tend to focus on non-relational properties. But
when two exemplars are present and the subject compares them, the subject will

focus on relational structure that was not noticed prior to the comparison. This predicts that comparison of representational structures encoding a causal relationship between desire-representation and behavior, for example comparison of (a) and (b), will enable children to notice and subsequently abstract a more general representation of the causal relationship between *mental representation* and action. Prior to comparison, children's grasp of the causal relationship between mental representation and action was only implicit because it was conflated with a specific type of representation, namely desire.

This line of reasoning fits nicely with Tomasello's (1992) 'verb-island' hypothesis, according to which children initially understand linguistic constructions with respect to specific verbs. For example, children might understand the passive form with the verb *hit* but not with the verb *love*. Similarly, I am interpreting Bartsch and Wellman's empirical data as evidence that desire-belief psychologists understand the representation-action construction with respect to the verb *desire* but not the verb *believe*. Interestingly, Tomasello endorses Gentner's theory about the role of analogy in learning abstract constructions, remarking that 'this is of course exactly the kind of cognitive ability needed for children to create a verb island schema across different arrangements of object participants.' (Tomasello, 2000, p. 242) I am making the specific claim that the comparison of desire-action representations can allow children to progressively align a more general causal relationship between mental representation and action. Through the 'deconflation' of this causal relationship, children are able to extend the relational role of causing behavior to other object participants, such as fact-presenting 'belief' content.[10]

I want to be clear about the type of contribution I am claiming that alignment-based learning makes to children's development of the belief concept. I am not claiming that structural-comparison, by itself, is sufficient to instruct the belief concept. However, a generalized representation-action construction makes available to children the hypothesis that *other* representations *besides* desire

---

[10] Also relevant here is work by Jacques and Zelazo (2005a, 2005b) on the interrelationship between labeling, cognitive flexibility, and ToM development. Jacques and Zelazo argue that labeling can contribute to improved cognitive flexibility and increased performance on ToM tasks. Tasks that measure cognitive flexibility are those that require subjects to select one perspective over another, and 'adopting the incorrect perspective is easier or more natural' (Jacques and Zelazo, 2005b, p. 55). This broad definition counts at least four dozen types of tasks as measuring cognitive flexibility, including tasks that require selection of explicit relational concepts over entity concepts (see Jacques and Zelazo, 2005b, Table 3.1). Jacques and Zelazo (2005a) offer two explanations for why labeling increasing flexibility. On the 'Levels of Consciousness' model, labels encourage self-reflection, thereby allowing one's mental states to become targets for higher levels of consciousness. On a second model, the fact that labels are arbitrary signifiers allows thinkers to achieve representational distance, which in turn facilitates greater control over one's representations. The structural-alignment analysis of contrastives provided above offers an alternative explanation. One way for children to achieve greater cognitive flexibility in a domain is to undergo a relational shift in that domain. The relational shift occurs as a result of increased domain knowledge, and language-induced structural-comparison is a primary mechanism for acquiring new relational domain knowledge.

play the relational role of causing behavior; it makes available the *hypothesis* that the fact-presenting 'belief' state plays an important role in the causal production of behavior. These explicit relational hypotheses can then be evaluated probabilistically in light of new cognitive-behavioral evidence. In contrast, there is no learning mechanism internal to the probabilistic models approach that is similarly well-suited to explain how children discover which mentalistic hypotheses to bring to bear for explicit, Bayesian testing. This contrast can be made clear by considering three challenges—two of them empirical, one conceptual—for current accounts of hypothesis search and discovery.

The first challenge concerns how children can know in advance which hypothesis to test. With respect to the belief concept, the challenge is to explain how it occurs to the child in the first place that belief-content might causally explain behavior. This challenge assumes that the representational resources for this hypothesis are available, and it asks how child are able to find—or antecedently appreciate the relevance of—these resources. As Gopnik and Wellman (2012) note, 'since an extremely large number of alternative hypotheses might be compatible with the evidence, we cannot learn by simply enumerating all the alternatives and testing each one' (p. 1097). Thus, the larger the search space, the bigger the challenge. In fact, I think that this problem is worse than many researchers realize. As discussed earlier, prior to their relational shift in a domain, children are biased in favor of attributes and against relations. This suggests that hypothesis search is more likely directed at attributes than at the systematic relations that need to figure in explanatory theories.[11] If finding the right hypothesis to test is already like finding a needle in a haystack, then being on the wrong side of the relational shift means that one is biased to search in all the wrong parts of the stack. Returning to the belief concept, the worry is that children will be more apt to consider the importance of attributional and non-systematic relational rather than systematic relational features of psychological categories.

Structure-mapping theory is well-suited to address this challenge because it describes general learning mechanisms that can naturally constrain hypothesis search. According to structure-mapping theory, representational alignment induces a focus specifically on relations. Moreover, alignment favors systematic relations over isolated relations. This suggests that comparison can direct subsequent hypothesis search to a particular sub-set of representational structures.[12] While I do not rule out that there can be other general learning mechanisms that constrain hypothesis

---

[11] Strong evidence for this claim can be found in research on relational transfer (Gick and Holyoak, 1983; Ross, 1989; Gentner, Rattermann and Forbus, 1993). This research indicates that surface representational elements dominate over structural features during retrieval (except in the case of experts).

[12] Referencing their study on comparison, Christie and Gentner (2010) note that 'the clearest reading of the results is that the children did not entertain the relational hypothesis in any real sense until they engaged in comparison' (p. 369). This inference is also compatible with the deconflation-effect of comparison as discussed in the context of the second challenge.

search,[13] it is noteworthy that structure-mapping theory is an already established and widely accepted research program that describes domain–general learning mechanisms well–tuned for this very challenge.

A second and related challenge is that the theoretically valuable hypotheses—those that are explicit about causal and functional relations—are in an important respect not available for children as search targets. Prior to their relational shift in a domain, children have conflated relations with particular object participants. For such implicit relational knowledge, the relational predicates needed for important explanatory and predictive work are 'stuck'—they are fused into representations of objects and contexts. Some process is needed to make these relational predicates accessible so that they can be manipulated as explicit hypotheses. Progressive–alignment through structural–comparison is precisely the type of learning that can allow children to meet this challenge. This is because it enables learners to start with their conflated representations, the comparison of which can then highlight and make available for abstraction theoretically valuable relations. The discussion of contrastives and children's comparison of desire–action representations demonstrates how this process might work for the formation of hypotheses about the causal efficacy of belief–content.[14]

The third challenge recalls Fodor's concept-learning argument and it has an important conceptual component. I argued earlier that if the Bayesian appeal to a

---

[13] Researchers working within the framework of the probabilistic models approach are pursuing possible solutions to this problem. Gopnik and Wellman (2012) describe what they regard as three promising approaches for addressing the search problem: children act in a way that involves 'informative experiments', children observe the actions of others, and children use sampling techniques. Gopnik and Wellman also discuss the potential of hierarchical Bayes nets, though they note that this approach may not bear on the production of new hypotheses (p. 1103). Putting aside concerns raised in Section 2 about the failure to distinguish between contexts of theory–failure and contexts of discovery, I want to emphasize that the current proposal describes specific cognitive mechanisms (i.e. structural alignment) that explain *how it is* that acting in the world and observing others acting in the world can constrain hypothesis search and make available new hypotheses.

[14] Gopnik and Wellman (2012) are skeptical of the role of analogical learning in children's development of new concepts. However, they do not consider the type of analogical learning that is described and applied in the current account. Gopnik and Wellman (p. 1103) offer two objections to analogy–based explanation of transformative, conceptual learning. First, they state that 'In order to recognize that a linguistic structure encodes some new, relevant conceptual insight, it seems that you must already have the conceptual resources that the structure is supposed to induce.' However, this objection either addresses just projective analogies (which I described earlier as a red herring in debates over conceptual discovery), or it does not consider the respect in which the sought after conceptual resources are present but implicit and conflated as argued in this essay. Second, the authors report that 'philosophers have long pointed out that the problem with analogical reasoning is the proliferation of possible analogies. Because an essentially infinite number of analogies are possible in any one case, how do you pick analogies that reshape your conceptual understanding in relevant ways and not get lost among those that will simply be dead ends or worse?' (p. 1103) However, this objection either addresses just projective analogies, or it fails to consider the relational and systematic–relational biases that constrain analogical processing during progressive alignment and schema abstraction.

distinction between latent and explicit hypotheses does not grant that there is extensive and innate propositional knowledge, then it leaves unanswered Fodor's target developmental question. Fodor's challenge now becomes: 'how can one convert "latent" hypotheses into "explicit" hypotheses in a way that does not involve explicitly hypothesizing about the latent hypotheses and which still counts as learning?' Some advocates of Bayesian modeling now appeal to a 'second type of manipulation' that performs the job of 'searching the latent hypothesis space and identifying the hypotheses that will become explicit hypotheses … what is being manipulated is not the hypotheses themselves but the primitives that define those hypotheses in the first place' (Perfors, 2012, p. 132). But this just reshuffles Fodor's objection. In order to address Fodor's objection, one would need to model or describe this second type of manipulation, and the described process would intuitively have to count as learning while not involving explicit hypothesis testing. On one reading, Fodor's point is simply that there are no (or cannot be) plausible candidates for such a process.

Structure-mapping processes, because they do not rely on hypothesis testing, are a promising candidate. The learning-role of 'feedback' will help clarify this point. In Fodor's examples of concept learning (Fodor, 1975, 1981), test subjects revise and confirm hypotheses on the basis of feedback they receive about the accuracy of their hypotheses. However, because the effectiveness of feedback presupposes a candidate knowledge structure (the hypothesis) that can be judged in reference to it, the learner travels in an epistemic circle. In contrast, the structural-alignment description of learning does not require learner feedback. Comparison activities lead to conceptual change and improve task performance without providing feedback to test subjects (see, e.g., Kotovsky and Gentner, 1996). In terms of the current proposal, the child does not need to hypothesize explicitly about the value of considering a more general causal relationship between mental content and behavior; that is not how this idea becomes an explicit hypothesis.[15] Rather, the symbolic or spatiotemporal juxtaposition of exemplars triggers the alignment of mental representations (not themselves hypotheses) with conflated relational predicates, which in turn engages the learning mechanisms of structural alignment (mechanisms that do not employ hypothesis formation and testing) so as to make available for explicit manipulation relational hypotheses that were previously not available for explicit manipulation.

Do these considerations solve Fodor's puzzle about concept learning? If we accept Fodor's claim that learning must be a rational process involving experiences that bear a 'confirmation relation' to hypotheses (Fodor, 1981), then the answer is 'no'. I assume that for Fodor, structural-alignment would not count as learning precisely *because* it does not require hypothesis testing. Instead it would be a 'brute-causal' process that facilitates concept 'acquisition.' While I

---

[15] In the language of hypothesis spaces, the child does not need to form explicit hypotheses about which latent 'hypotheses' to convert into explicit hypotheses.

am happy to grant that structural-alignment may not count as a mechanism of rational learning, I reject any identification of 'learning' with 'rational learning', and thus I reject any identification of 'learning mechanism' with 'rational learning mechanism'. This (false) identification likely stems from a (false) dichotomy between brute-causal and rational processes of conceptual change. Consider Fodor's claim that:

> [The] internal connection between concept learning and epistemic notions like *evidence* is the source of the strong intuition that concept learning is some sort of rational process. It contrasts sharply with kinds of concept acquisition where, for example, a concept is acquired by surgical implantation; or by swallowing a pill; or by hitting one's head against a hard surface, etc. Intuitively, none of these is concept *learning*; but any or all of them might eventuate in concept attainment (Fodor, 2008, p. 135).

However, these examples draw the intended contrast too sharply. Alignment-based learning, which is non-rational in Fodor's sense, is also not anything like having a coconut dropped on one's head and then having a new concept. Rather, alignment-based learning is a selected-for cognitive mechanism (Gentner, 2003) that causes the restructuring of conceptual knowledge. In addition, language and various other cultural tools routinely structure comparison opportunities in order to harness the potential of this learning mechanism. If learning through structural-alignment is like having a coconut dropped on one's head, then natural selection has designed us so that we process the effects of such strikes for epistemic gain, and culture has mechanized this process by setting up factories that implement just the right kind of coconut drops.

Nor is the relation between the input and the output of representational alignment *arbitrary*.[16] This idea is best understood in the context of Fodor's view that nearly all lexical concepts are unstructured. Fodor motivates this view by arguing that current theories of concepts (e.g. prototype theory) posit conceptual components (e.g. prototypes) that could not compose to form complex concepts. However, Fodor does not consider structure-mapping accounts of conceptual construction and restructuring. According to structure-mapping theory, complex concepts can emerge through forms of relational selectivity. This alternative compositional logic is implicit in the algorithms of computational models that simulate structural alignment (SME) (Falkenhainer, Forbus and Gentner, 1989) and progressive alignment (SEQL) (Kuehne, Forbus *et al.*, 2000). The relevance of these computational models

---

[16]   According to Fodor ' … one of the distinguishing characteristics of concept learning is the *nonarbitrariness* of the relation between what is learned and the character of the experiences that occasion the learning. (Compare the case of acquiring Latin by taking pills)' (Fodor, 1975, p. 37).

to Fodor's claims about structured concepts may be a promising topic for future research.[17]

4.3 Spatio–Temporal Juxtaposition of Cognitive-Behavioral Relations

In addition to symbolic juxtapositions, spatio-temporal juxtapositions provide a further opportunity for learners to compare and generate explicit relational knowledge. This section advances the conditional claim that if pretend role-play involves simulation, then pretend role-play serves as a spatio-temporal juxtaposition of cognitive-behavioral relations that enables children to progressively align the belief-desire schema. If correct, this application of structure-mapping principles has the additional benefit of analyzing simulative and theoretical heuristics as developmental complements rather than competitors. Pretense and simulation are complex topics, so the following is only a sketch.

A number of researchers argue that pretend role-play involves a simulation heuristic in which the pretender disengages from their current perspective in order to simulate the perspective of someone else (see Gordon, 1992; Harris, 2000; Lillard, 2001; Goldman, 2006). For example, pretending that a figurine is a dancer would involve simulating a dancer and then projecting one's own mental states and intentions onto the figurine. Or, pretending to be a dancer oneself involves simulating the mental states of a dancer and, to a certain extent, having these mental states influence one's behavior.

My suggestion is that simulative role-play provides children an opportunity to align cognitive-behavioral representations, and thus a further opportunity to make explicit the relations that define belief-desire psychology. This suggestion finds support in an empirical model of simulation as a cognitive process that recruits structural alignment and facilitates the development of folk-psychological theory (see Bach, 2011, for a description of this model). The basic prediction of this model is that running an offline simulation and then projecting the results of that simulation onto a target will require, often at several junctures, aligning cognitive-behavioral representations (e.g. a representation of oneself and the target). If role-play is simulative, this view suggests how role-play can lead to relational abstractions that inform the

---

[17] Carey's (2009) 'Quinean bootstrapping' account also emphasizes a route to greater representational expressive power that does not depend on hypothesis-testing. While Carey also provides a role for analogy in her account, that role is very different from the role given to analogical cognition in the current proposal. The current proposal emphasizes how the comparison of attributionally similar but meaningful representations can highlight relational features, thereby making these relations explicit targets for, e.g., probabilistic reasoning. In contrast, Carey emphasizes the role of projective analogical mapping in children's interpretation of public symbols that are semantically meaningless (at least in terms of wide-content) place-holders. For this reason, criticisms of the role of analogy in transformative learning that take Carey's account as standard—for example (Rey, 2011) and (Gopnik and Wellman, 2012)—will likely fail to address the merits of the current proposal. See also the discussion in Note 14.

belief-desire schema. During role-play, the spatio-temporal juxtaposition of oneself and the object of role play invites representational alignment and thereby supports relational abstraction.

Consider a three year who pretends that a doll 'wants her mommy' (Harris, 2000), and then brings the doll to 'mommy's location'. According to the simulation construal of role-play, this child will project her own desire for mommy as well as her behavioral intention onto the doll. Three-year-olds understand the relationship between cognition and behavior according to their current psychological theory, for example, desire-belief psychology. As such, the projected action schema will be encoded by the desire-belief theory, and thus the pretending child finds herself aligning instances of the desire-belief schema. Moreover, pretending children plausibly toggle back in forth between their own perspective, where they recruit causal knowledge, to the pretend world of the doll, where this knowledge is imported (Harris, 2000). Such toggling multiplies the opportunity for representational alignment. The thesis of progressive alignment through structural-comparison predicts that repeated experience in the alignment of instantiations of action-schemas during role-play will make explicit the deeper and more explanatory cognitive-behavioral structures that define the belief-desire schema.[18]

4.4 Acquiring a Representational Theory of Mind

The previous sections targeted the transition from desire-belief psychology to belief-desire psychology. Earlier developmental progressions, such as the shift from a non-representational to a representational theory of mental states, are open to a similar structure-mapping analysis. There appear to be at least two domains—pictures and language—in which children understand representational relationships prior to when they understand them in the cognitive domain. I discuss each domain in turn, examining how comparison opportunities can facilitate the discovery of mentalistic hypotheses.

Children initially try to pick up two-dimensional pictured objects, but eventually they discover that the contents of pictures are representational rather than real. For example, nine-month-olds try to pick up pictured objects as if they were real (DeLoache *et al.*, 1998), but 'by 19 months, infants instead point at and label pictures, indicating a nascent awareness of how pictures differ from the real entities they represent' (Simcock and DeLoache, 2006, p. 1352). This suggests that the epistemic transition in the domain of pictures occurs prior to infant's grasp that minds contain representations. However, through the repeated comparison of pictures, children could abstract a more relationally explicit understanding of the representing relation as this relation obtains in the domain of pictures and artifacts. Consider children who are able to compare two attributionally similar photos—for

---

[18] Pretense-based spatio-temporal juxtapositions are also culturally invited. Caregivers use various tools and particularly language to encourage pretend play, e.g. 'let's pretend that we are dancers'.

example, two pictures of familiar objects on the same page of a children's book. With repeated comparison, children can develop a more explicit understanding of the representing relation so that they can extend it to, and now make comparisons between, a greater variety of pictures (and artifacts) with representational content. Indeed, as the process of progressive alignment continues, children may eventually refine their grasp of the representing relation to the point where they can extend it to object participants in less familiar domains, for example the cognitive domain. The continuing deconflation of the representing-represented relation can help make available to children (specifically, the desire-psychologist) the hypothesis that desires, understood initially as just internal drives of some sort, are object participants in the representation–represented relation.

With respect to language, de Villiers (2000) points out that children grasp that the embedded *says-that* complement can be false well before they grasp that the embedded *think-that* complement can be false. de Villiers claims that grasp of the *says-that* complement structure bootstraps grasp of the *think-that* complement structure because *says-that* can provide overt evidence of a difference in truth–value between the embedded clause and the sentence. In de Villiers' terminology, it is easier for children to grasp that the embedded clause of *says-that* constructions refers to someone's *point-of-view*. A structural-alignment analysis provides further detail for how exposure to *says-that* constructions can make 'point-of-view'—a relation between thinkers and propositions—relationally explicit so that it can be predicated to the cognitive domain (*think-that*). According to this analysis, the bootstrapping effect is most likely to obtain when *says-that* constructions are juxtaposed and compared (e.g. 'Mommy said that *x*, and daddy said that *y*'). Moreover, progressive alignment of the point-of-view relation would not require disparity in truth-values between the embedded clause and sentence of either of the aligned representations.

Recalling Tomasello's terminology, we can summarize by claiming that children's understanding of the representation-represented construction is initially restricted to the 'island' of observable artifacts and the 'island' of says-that complements, and that comparisons between exemplars in these familiar islands will foster the abstraction of a more general relational construction that can then be transferred to the cognitive domain.

5. Implications of a Precocious ToM

The developmental timetables from direct false-belief task studies have been increasingly challenged. Data from studies that measure ToM development using spontaneous response tasks suggest that infants in the second year of life can attribute, and make behavioral predictions on the basis of, false beliefs (Baillargeon, Scott and He, 2010).[19] Several researchers have offered alternative, less mentalistic

---

[19] Consistent with this view, several researchers claim that it is children's underdeveloped executive function skills and not a conceptual deficit that explains the results from direct false-belief

explanations of these data in terms of children's use of behavioral rules (Perner, 2010), knowledge of awareness versus unawareness (Wellman, 2011), and employing a 'teleological stance' (Zadwidzki, 2011). Nonetheless, it is worthwhile to consider what implications follow for the proposed account if young infants have a precocious grasp of false-belief. If mentalistic interpretations of data from spontaneous false-belief tasks are correct, I see two possible roles for alignment-based learning in children's development of psychological concepts. One of these roles makes a fundamental concession to nativism and the other does not. I examine each in turn.

Throughout I have used the term 'implicit' to mean relational knowledge that is contextually embedded and conflated with specific argument values, and 'explicit' to mean relational knowledge that is desituated and subserved by a flexible relational predicate not conflated with particular argument values. In order to sketch the first application I need to make another, albeit familiar, representational distinction between *tacit* and *salient* knowledge structures (however, I suspect that this distinction is often conflated with the difference between explicit and implicit knowledge as defined here). Salient representations are conscious and available for verbal report, while tacit representations are not fully conscious and not available for verbal report.[20] While the two distinctions can correspond (implicit knowledge that is tacit; explicit knowledge that is salient), they often come apart. For example, in the debate over the ontology of mental states, metaphysical functionalists accuse identity theorists of conflating mental states with particular object-realizers, namely, brain states. If these critics are correct, then assuming identity-theorists are conscious of and can verbalize their own positions, this knowledge would be both salient and implicit. An example of tacit, explicit knowledge may be folk-linguistic knowledge of prepositions.

The first application assumes that spontaneous false-belief tasks reveal children's tacit and relationally explicit knowledge of the relational-role categories *belief* and *desire*. The role of structural-comparison would be to convert this tacit, relationally explicit knowledge into salient, relationally explicit knowledge. The aligned representations would be tacit, but the results of such alignment would be relational representations that are conscious and verbalizable. On this construal, the types of structural alignment discussed in Sections 4.2 and 4.3 would facilitate the salient discovery (or rediscovery, if you prefer) of relational hypotheses. The conversion of specifically relational predicates from tacit to salient representational format would be very useful if certain processing and theory-building mechanisms (for example, probabilistic causal modeling) require or work best with salient representations.

---

tasks. The lack of executive function skills creates performance errors during the direct false-belief task.

[20] The vigorous debates between simulation-theorists and theory-theorists provide ample discussion of where and how to draw this distinction. See especially Stich and Nichols, 1992; Heal, 1994; Botterill, 1996; Perner, 1999; Davies and Stone, 2001; Goldman, 2006.

Note that this reconstrual of the role of alignment-based learning in ToM development makes a fundamental concession to nativism. This is because it does not offer a domain-general explanation of the development of tacit representations of *belief* and *desire*.[21]

The second application speculatively extends the structure–mapping explanations of Sections 4.2 and 4.3 into earlier periods of development in order to explain the origination of tacit, relationally explicit representation of the categories *belief* and *desire*. It is important to note that the proposals of Sections 4.2 and 4.3 are applications of a more general schema. According to this schema, the relational structures that develop through structural-alignment are always constrained by existing domain knowledge (Section 3). With respect to a precocious ToM, the question becomes: what type of domain knowledge could inform structural-alignment in order to facilitate the abstraction of a (tacitly understood) belief–desire schema at, say, 15 months?

Kuehne, Gentner, and Forbus (2000) attribute alignment-based rule-learning to 7 month-old infants. If this is correct, then structural-alignment is available as a learning mechanism for tacit ToM. Future research would need to investigate how comparison opportunities might interact with cognitive-behavioral domain knowledge at this age. One fruitful possibility is that the activity of *joint attention* functions to promote comparison and facilitate abstraction of cognitive-behavioral relations. Joint attention is an interesting candidate because it appears to be developmentally prior to early ToM, occurring at 9 months and as early as 5 months (Striano and Bertin, 2005), and also because it involves a triadic exchange between infant, object, and another person. It is a feature of this triadic exchange that the infant is in some sense aware that the other is aware of the object (Hobson, 2005).[22] The speculation on offer is that infants juxtapose and align representations that encode subjective orientations (their own and the other's) towards the triangulating object during joint attention. If this is correct, then such comparisons could make explicit the shared attentional relation, after which it could be incorporated and evaluated in a cognitive-behavioral theory.

---

[21] This reconstrual also bears similarities to (Karmiloff-Smith, 1992), which claims that procedural knowledge can become salient through an iterative, spontaneous, and internally driven process of representational redescription. According to Karmiloff-Smith, once knowledge structures are redescribed in the new representational format they are available for conscious access, verbal report, and theory-building. Structure-mapping descriptions of representational change are similar to Karmiloff-Smith's proposal but with greater focus on mechanism, i.e. progressive alignment and relational re-representation (see Gentner *et al.*, 1995). Important differences between the accounts are that structure-mapping principles allow for representational redescription to occur early and piece-meal (prior to what Karmiloff-Smith terms behavioral mastery) and as a result of external prompts (e.g. cultural invitations for comparison) (Gentner *et al.*, 1995).

[22] See the chapters in Elian *et al.*, 2005, for discussions of the extent to which jointly attending infants understand attention as a mental state. See also the research by Tomasello and colleagues (e.g. Tomasello *et al.*, 2005) for a general discussion of how shared intentionality and joint attention can facilitate the development of folk-psychological skill.

6. Conclusion

Accounts that explain ToM development in terms of domain-general transitional mechanisms fail to explain children's discovery of mentalistic hypotheses. This essay proposed that the domain-general learning mechanism of progressive alignment through structural-comparison indicates how children discover mentalistic hypotheses that can then be theoretically evaluated. In arguing for this thesis, I claimed that progressive alignment through structural-comparison explains how children convert implicit relational knowledge into explicit relational knowledge in order to develop the relational knowledge structures that define belief-desire psychology. I described how contrastive conversations about mental states serve as symbolic juxtapositions of cognitive-behavioral causal relationships, and suggested that simulative role-play serves as a spatio-temporal juxtaposition of cognitive-behavioral relationships. Both forms of juxtaposition induce structural-alignment and facilitate abstraction of relational constructions that correspond to the essential relations of belief-desire psychology. I further suggested that earlier achievements, such as a representational theory of mind, or a precocious and tacit ToM, are subject to a similar structure-mapping analysis. In addition, I claimed that the proposed view strikes a middle ground with respect to Fodor's puzzle about concept learning.

If successful, the current proposal indicates several new avenues for theoretical development and empirical research. I conclude by mentioning three. First, the current proposal has a number of testable implications, several of which are made explicit in the essay (for example, empirical predictions about the effects of contrastives and picture comparison). Second, by modeling an important role for a distinct general learning mechanism—structural-alignment—in ToM development, the current proposal indicates new connections to explore with respect to the cognitive style of autism. For example, if it can be shown that certain non-mentalistic diagnostic features of autism—particularly weak generativity (Jarrold, 1997; Peterson and Bowler, 2000)—stem from an impaired ability for analogical reminding and relational retrieval, then the cognitive mechanisms that subserve analogical and similarity cognition could offer resources for a more unified analysis of autism than is provided by extant accounts that tend to explain either just mentalistic features or just non-mentalistic features. Third, the current proposal might be generalized in two directions. First, the proposal codifies a general schema for cognitive development that involves the iterated and dialectical use of different domain-general learning mechanisms toward the deconflation and explicitization of relational predicates. This schema could be generalized to other domains—for example the biological and the physical—that nativists target with challenges analogous to those considered here. Second, if the problem of conflated relations occurs at all levels of development, then the current account might generalize to other periods of development within a single domain. In particular, recent work on uniform relational encoding and analogical reminding suggests

an important role for structural alignment in the explanation of the development of expertise.

Department of Philosophy
Bowling Green State University
Firelands College

References

Bach, T. 2011: Structure-mapping: directions from simulation to theory. *Philosophical Psychology*, 24, 23−51.

Bach, T. 2012: Analogical cognition: applications in epistemology and the philosophy of mind and language. *Philosophy Compass*, 7(5), 348−60.

Baillargeon, R., Scott, S. M. and He, Z. 2010: False-belief understanding in infants. *Trends in Cognitive Science*, 14, 110−18.

Baldwin, D. A. and Saylor, M.M. 2005: Language promotes structural alignment in the acquisition of mentalistic concept. In J. W. Astington and J. A. Baird (eds), *Why Language Matters for Theory of Mind*. New York: Oxford University Press.

Bartsch, K. and Wellman, H.M. 1995: *Children Talk about the Mind*. New York: Oxford University Press.

Boroditsky, L. and Ramscar, M. 2001: 'First, we assume a spherical cow … '. *Behavioral and Brain Sciences,* 24, 656−7.

Botterill, G. 1996: Folk psychology and theoretical status. In P. Carruthers and P. K. Smith (eds), *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Carey, S. 2009: *The Origin of Concepts*. Oxford: Oxford University Press.

Christie, S. and Gentner, D. 2010: Where hypotheses come from: learning new relations by structural-alignment. *Journal of Cognition and Development*, 11(3), 356−73.

Davies, M. and Stone, T. 2001: Mental simulation, tacit theory, and the threat of collapse. *Philosophical Topics,* 29, 127−74.

DeLoache, J. S., Pierroutsakos, S. L., Uttal, D. H., Rosengren, K. S. and Gottlieb, A. 1998: Grasping the nature of pictures. *Psychological Science*, 9, 205−10.

de Villiers, J. 2000: Language and theory of mind: what are the developmental relationships? In S. Baron-Cohen, H. Tager-Flusberg and D. Cohen (eds), *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*, 2nd edn. Oxford: Oxford University Press.

Elian, N., Hoerl, C., McCormack, T. and Roessler, J. (eds) 2005: *Joint Attention: Communication and Other Minds*. New York: Oxford University Press.

Falkenhainer, B., Forbus, K. D. and Gentner, D. 1989: The structure-mapping engine: algorithm and examples. *Artificial Intelligence*, 41, 1−63.

Fisher, C. 2000: From form to meaning: a role for structural analogy in the acquisition of language. In H. W. Reese (ed.), *Advances in Child Development and Behavior*. New York: Academic Press.

Fodor, J. 1975: *The Language of Thought*. Cambridge, MA: Harvard University Press.

Fodor, J. 1980: Fixation of belief and concept acquisition. In M. Piatelli-Palmarini (ed.), *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*. Cambridge, MA: Harvard University Press.

Fodor, J. 1981: The present status of the innateness controversy. In his *Representations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, MA: MIT Press.

Fodor, J. 2008: *LOT 2: The Language of Thought Revisited*. Oxford: Oxford University Press.

Forbus, K. D., Gentner, D., Markman, A. B. and Ferguson, R. W. 1998: Analogy just looks like high level perception: why a domain general approach to analogical mapping is right. *Journal of Experimental and Theoretical Artificial Intelligence*, 10, 231–57.

Gentner, D. 2003: Why we're so smart. In D. Gentner and S. Goldin-Meadow (eds), *Language in Mind: Advances in the Study of Language and Thought*. Cambridge, MA: MIT Press.

Gentner, D. 2005: The development of relational category knowledge. In L. Gershkoff-Stowe and D. H. Rakison (eds), *Building Object Categories in Developmental Time*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gentner, D. and Colhoun, J. 2010: Analogical processes in human thinking and learning. In B. Glatzeder, V. Goel and A. von Müller (eds), *On Thinking: Vol. 2. Towards a Theory of Thinking*. Berlin: Springer Verlag.

Gentner, D. and Forbus, K. 2011: Computational models of analogy. *WIREs Cognitive Science*, 2, 266–76.

Gentner, D. and Kurtz, K. 2005: Relational categories. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman and P. W. Wolff (eds), *Categorization Inside and Outside the Lab*. Washington, DC: American Psychological Association.

Gentner, D. and Namy, L. 1999: Comparison in the development of categories. *Cognitive Development*, 14, 487–513.

Gentner, D and Namy, L. 2004: The role of comparison in children's early word learning. In D. G. Hall and S. R. Waxman (eds), *Weaving a Lexicon*. Cambridge, MA: MIT Press.

Gentner, D., Rattermann, M. J. and Forbus, K. D. 1993: The roles of similarity in transfer: separating retrievability from inferential soundness. *Cognitive Psychology, 25,* 524–75.

Gentner, D., Rattermann, M., Markman, A. and Kotovsky, L. 1995: Two forces in the development of relational similarity. In T. J. Simon and G. S. Halford (eds), *Developing Cognitive Competence: New Approaches to Process Modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gentner, D., Brem, S., Ferguson, R., Wolff, P., Markman, A. B. and Forbus, K. D. 1997: Analogy in discovery: a case study of Johannes Kepler. In T. B. Ward, S. M. Smith and J. Vaid (eds), *Creative Thought: An Investigation of Conceptual Structures and Processes*. Washington, DC: American Psychological Association.

Gick, M. L. and Holyoak, K. J. 1980: Analogical problem solving. *Cognitive Psychology,* 12, 306–55.

Gick, M. L. and Holyoak, K. J. 1983: Schema induction and analogical transfer. *Cognitive Psychology*, 15, 1–38.

Goldman, A. 2006: *Simulating Minds*. Oxford: Oxford University Press.

Gopnik, A. 2011: A unified account of abstract structure and conceptual change: probabilistic models and early learning mechanisms. Commentary on Susan Carey 'The Origin of Concepts'. *Behavioral and Brain Sciences*, 34, 129–30.

Gopnik, A. and Meltzoff, A. 1997: *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.

Gopnik, A. and Schulz, L. E. 2004: Mechanisms of theory-formation in young children. *Trends in Cognitive Science*, 8, 371–7.

Gopnik, A. and Wellman, H. 2012: Reconstructing constructivism: causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, 138(6), 1085–1108.

Gordon, R. 1992: The simulation theory: objections and misconceptions. *Mind & Language*, 7, 158–71.

Harris, P. 2000: *The Work of the Imagination*. Oxford: Blackwell.

Heal, J. 1994: Simulation vs. theory theory: what is at issue? In C. Peacocke (ed), *Objectivity, Simulation, and the Unity of Consciousness*. Oxford: Oxford University Press.

Hobson, R. P. 2005: What puts the jointness into joint attention? In N. Elian, C. Hoerl, T. McCormack and J. Roessler (eds), *Joint Attention: Communication and Other Minds*. Oxford: Oxford University Press.

Jacques, S. and Zelazo, P. D. 2005a: Language and the development of cognitive flexibility: implications for theory of mind. In J. W. Astington and J. A. Baird (eds), *Why Language Matters for Theory of Mind*. New York: Oxford University Press.

Jacques, S. and Zelazo, P. D. 2005b: On the possible roots of cognitive flexibility. In B. D. Homer and C. S. Tamis-Lemonda (eds), *The Development of Social Understanding and Communication*. Mahwah, NJ: Lawrence Erlbaum Associates.

Jarrold, C. 1997: Pretend play in autism: executive explanations. In J. Russell (ed.), *Autism as an Executive Disorder*. New York: Oxford University Press.

Karmiloff-Smith, A. 1992: *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: MIT Press.

Kotovsky, L. and Gentner, D. 1996: Comparison and categorization in the development of relational similarity. *Child Development*. 67, 2797–2822.

Kuehne, S. E., Gentner, D. and Forbus, K. D. 2000a: Modeling infant learning via symbolic structural alignment. *Proceedings of the Twenty-second Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates, 286–91.

Kuehne, S.E., Forbus, K.D., Gentner, D. and Quinn, B. 2000b: SEQL: category learning as incremental abstraction using structure mapping, *Proceedings of the Twenty-second Annual Conference of the Cognitive Science Society.*

Leslie, A., Friedman, O. and German, T. 2004: Core mechanisms in 'theory of mind'. *Trends in Cognitive Science*, 8, 528–33.a

Lillard, A. 2001: Pretend play as twin earth: a social-cognitive analysis. *Developmental Review*, 21, 495–531.

Lowenstein, J., Thompson, L. and Gentner, D. 1999: Analogical encoding facilitates knowledge transfer in negotiation. *Psychonomic Bulletin and Review*, 6(4), 586–97.

Markman, A. and Stillwell, C. 2001: Role-governed categories. *Journal of Experimental & Theoretical Artificial Intelligence*, 13, 329–58.

Namy, L. and Gentner, D. 2002: Making a silk purse out of two sow's ears: young children's use of comparison in category learning. *Journal of Experimental Psychology*, 131, 5–15.

Perfors, A. 2012: Bayesian models of cognition: what's built in after all? *Philosophy Compass*, 7(2), 127–38.

Perfors, A., Tenenbaum, J. B., Griffiths, T. L. and Xu, F. 2011: A tutorial introduction to Bayesian models of cognitive development. *Cognition*, 120, 302–21.

Perner, J. 1999: Simulation as explication of prediction-implicit knowledge about the mind: arguments for a simulation-theory mix. In P. Carruthers and P. K. Smith (eds), *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Perner, J. 2010: Who took the cog out of cognitive science? Mentalism in an era of anti-cognitivism. In P. A. Frensch and R. Schwarzer (eds), *Cognition and Neuropsychology: International Perspectives on Psychological Science (Volume 1)*. Hove: Psychology Press.

Peterson, D. and Bowler, D. M. 2000: Counterfactual reasoning and false belief understanding in children with autism. *Autism*, 4(4), 391–405.

Rey, G. 2011: Learning, expressive power and mad dog nativism: the poverty of stimuli (and analogies), yet again. *Presentation to the Joint Conference for the European and American Societies for Philosophy and Psychology, Montreal.*

Ross, B. H. 1989: Distinguishing types of superficial similarities: different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning,Memory and Cognition*, 15(3), 456–68.

Ryle, G. 1949: *The Concept of Mind*. Chicago, IL: University of Chicago Press.

Scholl, B. and Leslie, A. 1999: Modularity, development and 'theory of mind'. *Mind & Language*, 14, 131–53.

Schroeder, T. 2004: *Three Faces of Desire*. Oxford: Oxford University Press.

Simcock, G. and Deloache, J.S. 2006: Get the picture? The effects of iconicity on toddlers' re-enactment from picture. *Developmental Psychology, 42,* 1352−7.

Stich, S. and Nichols, S. 1992: Folk psychology: simulation or tacit theory? *Mind & Language*, 7, 35−71.

Striano, T. and Bertin, E. 2005: Coordinated affect with mothers and strangers: a longitudinal analysis of joint engagement between 5 and 9 months of age. *Cognition and Emotion*, 19(5), 781−90.

Tomasello, M. 1992: *First Verbs*. New York: Cambridge University Press.

Tomasello, M. 2000: Do young children have adult syntactic competence? *Cognition*, 74, 209−53.

Tomasello, M., Carpenter, M., Call, J., Behne, T. and Moll, H. 2005: Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences*, 28, 675−735.

Wellman, H. 2011: Developing a theory of mind. In U. Goswami (ed.), *The Blackwell Handbook of Childhood Cognitive Development*, 2nd edn. Oxford: Wiley-Blackwell.

Wellman, H. M., Cross, D. and Watson, J. 2001: Meta-analysis of theory of mind development: the truth about false-belief. *Child Development,* 72, 655−84.

Zadwidzki, T. W. 2011: How to interpret infant socio-cognitive competence. *Review of Philosophy and Psychology*, 2(3), 483−97.