Robert Mirski

Socio-cognitive development in context: Theory of mind, folk psychology, and culture from an interactivist perspective

Rozwój społeczno-poznawczy w kontekście: Teoria umysłu, psychologia potoczna i kultura z perspektywy interaktywistycznej

Rozprawa doktorska przygotowana
pod kierunkiem dr hab. Arkadiusza Guta
na Uniwersytecie Mikołaja Kopernika w Toruniu

Toruń, 2024

Acknowledgments

First and foremost, I would like to extend my thanks to my supervisor, Arkadiusz Gut, for teaching me the value of persistence and for his fatherly support at every step of the way. His insightful guidance and unwavering belief in my potential have been the cornerstone of my academic growth. I am especially grateful for his encouragement to aim higher, even in moments of doubt.

I am profoundly thankful to Mark Bickhard, whose visionary theoretical framework lies at the heart of this dissertation. Working with him personally was an extraordinary privilege and a defining moment in my intellectual development. His kindness, patience and openness to exploring my ideas, no matter how nascent, left an enduring impression on me and set a standard of intellectual generosity and mentorship that I aspire to emulate throughout my academic journey.

A heartfelt thanks goes to Jed Allen, who supported me in multiple ways during my first, tentative steps into the world of academia. I greatly value our intellectual exchanges and his insights, which played a significant role in shaping the account presented in this dissertation.

I also deeply appreciate the thoughtful discussions I had with Joanna Klara Teske, which have enriched my thinking far beyond the scope of this dissertation. Her sharp intellect and curiosity were a constant source of inspiration, and I truly value the lasting impact of our exchanges on my work and perspective.

I want to express my gratitude to Anna Karczmarczyk, who became an important part of my life toward the end of this journey. Her kindness, support, and understanding during these final stages have meant more to me than words can convey. Her presence has been a source of warmth and encouragement, making the challenges of completing this dissertation much more bearable. For her care and belief in me, I am deeply thankful.

With profound gratitude, I honor the memory of my dear friend, Mateusz Chról, whose kindness and eccentric intellect were a profound source of inspiration throughout the work on this project. His mind was a treasure trove of witty observations and unconventional ideas, the absence of which leaves a profound void that will be deeply felt and forever missed.

Finally, I would like to thank my parents for their constant support and for taking care of me when I was really sick. They gave me the strength to recover and rediscover my enthusiasm for both life and my academic journey.

Table of Contents

1. Theoretical context

This dissertation concerns the role of socio-cultural factors in the development of human social cognition. Social cognition is conceptualized in the broad sense as the ability to understand and competently interact with one's conspecifics. This includes the ability to attribute mental states to others and reason about their perspective, known in the literature as *theory of mind (ToM), mindreading* or *mentalization*, which has received considerable attention over the recent decades (Fenici, 2017; Wellman, 2018). There is a growing number of empirical results that seem to indicate that socio-cultural context greatly influences the child's ability to reason about other minds (Milligan, Astington, & La Dack, 2007; Slaughter & Perez-Zapata, 2014; Szpak & Białecka-Pikul, 2020; Taumoepeau, 2019; Tompkins, Benigno, Kiger Lee, & Wright, 2018). For instance, children in the West tend to understand false beliefs (as measured by the false-belief task) around age 4 while for the Pacific Islander cultures it is as late as 12 (Mayer & Träuble, 2012). This has created an interpretative challenge for the extant theoretical models of theory of mind and social cognition. In this dissertation, I offer a critique of the mainstream attempts to make sense of the observed cultural influence, and develop an alternative account applying the interactivist ontology of the mind to the problem (Bickhard, 2009).

The critical part of this dissertation challenges the ontological foundations of traditional Theory of Mind (ToM) models, arguing that their metaphysical assumptions are conceptually flawed and inadequate for explaining the enculturation of social cognition. Leveling such foundational criticism presupposes that metaphysical assumptions underlying scientific research programs can indeed be critically examined. This contrasts with positions, such as that of Lakatos (1970), which hold that research programs can compete only on the basis of empirical growth they generate rather than rational examination of their metaphysical "hard core". However, such assumptions, while not empirically testable, shape the formulation of hypotheses, the design of studies, and the generation and interpretation of data. Accordingly, the present work assumes that their coherence, compatibility with other knowledge, and practical implications can be systematically examined, making them legitimate subjects of purely theoretical critique.

The positive account of socio-cognitive development presented in this dissertation is most fundamentally rooted in the interactivist model. Interactivism, the lifetime work of Mark H. Bickhard, is an action-based model of cognition, social realities, and "whole persons"

(Bickhard, 2024). Drawing inspiration from American pragmatism, Piagetian constructivism, and evolutionary epistemology (D. T. Campbell, 1974), interactivism situates cognition as an emergent property of adaptive systems. It models mental representation and normativity as arising naturally from the self-maintaining dynamics of living systems. While sharing ground with enactivism (Bickhard, 2016) and ecological psychology (Bickhard & Richie, 1983), interactivism provides a distinct framework for understanding the ontological emergence of mental content without foundational representations. Within interactivism, social cognition arises from the progressive development of interactive conventions and reflective capacities, enabling children to navigate and shape complex social realities.

While the interactivist model remains the overarching ontological framework for the account developed in this dissertation, there are also other inspirations from a range of complementary theories in developmental psychology and cognitive science. Katherine Nelson's theory of cognitive development has played a pivotal role in my thinking as well, particularly her emphasis on the importance of language and social interaction. Her work underscores how cultural practices and linguistic interactions embed children in a shared social reality, forming a foundation for their understanding of others' minds. Similarly, Carpendale and Lewis' approach to social understanding, which highlights the co-construction of knowledge through social interaction and dialogue, provides valuable insights into how socio-cognitive abilities are scaffolded in relational contexts. Moreover, I have benefited greatly from reading the work of philosophers engaged in the debates on Theory of Mind, such as Marco Fenici's empirical argumentation for a socio-cultural interpretation of ToM data, much aligned with the account of this dissertation. In my argumentation against nativist positions, I have drawn extensively on discussions in evolutionary developmental biology (evo-devo), especially Susan Oyama's developmental systems theory and Stephen Jay Gould's critique of preformationist thinking, both of which emphasize the contingent, dynamic, and multifactorial nature of development as opposed to fixed, innate structures.

2. Significance of the dissertation

There are several novel contributions that this dissertation offers to the field. First, the project reviews and discusses empirical data on theory of mind development across cultures. There have been only a handful of similar reviews in the literature to date (Lavelle, 2019; Slaughter & Perez-Zapata, 2014; Taumoepeau, 2019), and the one included in the present work

extends and complements them in unique ways. Second, the dissertation offers a critique of the dominant theoretical models of theory of mind that is unprecedented within the area of research. It has two aspects: First, I take Bickhard's general criticism of what he has termed "encodingism" in theories of cognition and apply it specifically to the field of Theory of Mind. The main charge is the logically necessary but conceptually untenable stipulation of cognitive foundations from which further development of social cognition proceeds, which is termed *foundationalism*. Additionally, since the idea of cognitive foundations being innate is often taken as a way out of the problem of foundationalism, I argue that the concept of innateness is theoretically untenable as well. Third, the project advances a positive account of social cognition and theory of mind nested within the interactivist ontology. This adds a promising alternative to the existing accounts, one which offers multiple unique insights – e.g. it makes a case for close links between socio-cognitive abilities and social ontology. Fourth, I engage, to a varying extent, with other partly convergent models that are being developed today (e.g. enactivism, FEP). Finally, despite its unique insights, interactivism remains a largely ignored and often misunderstood theory of cognition – the presentation of the model included in the project humbly hopes to offer an accessible introduction to the theory.

3. Description of the articles

The dissertation is composed of five journal articles and two short commentaries. They are listed below together with abbreviations (in bold) I will use to refer to the papers throughout the summary of the main argument presented in section 5. The articles can be found at the end of this document, in section 7.

1. Mirski, R. (2018). Współczesne trudności i wyzwania w myśleniu o teorii umysłu: Perspektywa międzykulturowa. In A. Gut & Z. Wróblewski (Eds.), Filozofować eksperymentalnie: Umysł w świecie kultury (pp. 53–73). Wydawnictwo KUL. **WT**
2. Mirski, R., & Gut, A. (2020). Action-based versus cognitivist perspectives on socio-cognitive development: culture, language and social experience within the two paradigms. *Synthese*, 28(2), 96. https://doi.org/10.1007/s11229-018-01976-y **AvC**

3. Mirski, R., & Bickhard, M. H. (2021). Conventional minds: An interactivist perspective on social cognition and its enculturation. *New Ideas in Psychology*, 62, 100856. https://doi.org/10.1016/j.newideapsych.2021.100856 **CM**
4. Mirski, R. (2019). Krytyka natywizmu jawnego i ukrytego w badaniach nad dziecięcymi teoriami umysłu. *Psychologia Rozwojowa*, 24(2), 15–28. https://doi.org/10.4467/20843879PR.19.007.10890 **KN**
5. Allen, J. W. P., Mirski, R., & Bickhard, M. H. (2024). Beyond the mirror: an action-based model of knowing through reflection. *Frontiers in Developmental Psychology*, 2, Article 1449705. https://doi.org/10.3389/fdpys.2024.1449705 **BM**

6. Mirski, R., & Bickhard, M. H. (2019). Encodingism is not just a bad metaphor. *Behavioral and Brain Sciences*, 42, e237. https://doi.org/10.1017/S0140525X19001286 **EM**
7. Mirski, R., Bickhard, M. H., Eck, D., & Gut, A. (2020). Encultured minds, not error reduction minds. *Behavioral and Brain Sciences*, 43, e109. https://doi.org/10.1017/S0140525X19002826 **ERM**

The first paper on the list, "**Współczesne trudności i wyzwania w myśleniu o teorii umysłu: Perspektywa międzykulturowa**" (WT), acts as a starting point for exploring the key issues discussed in the dissertation. It examines the three dominant models of Theory of Mind (ToM)—the modular nativist, rational constructivist, and two-systems approaches—within the context of cross-cultural empirical data. This paper highlights significant challenges in how these models interpret cross-cultural variability, arguing that their theoretical frameworks lack the conceptual tools to account for such diversity. By situating these models against the backdrop of cross-cultural studies, this paper sets the stage for the foundational criticisms and alternative proposals developed in subsequent texts.

The second paper, "**Action-based versus cognitivist perspectives on socio-cognitive development: culture, language and social experience within the two paradigms**" (AvC), expands the scope of the critique introduced in WT. It delves deeper into the limitations of traditional ToM frameworks, focusing on their ontological assumptions. Drawing on Mark Bickhard's critique of encodingism, this paper challenges the foundational premises of cognitivist perspectives, arguing that they impose restrictive conceptual constraints on the

understanding of socio-cognitive development. Beyond critique, AvC introduces the interactivist framework as a promising alternative, offering a preliminary interpretation of the cross-cultural empirical data. This paper bridges the discussion between the problems of existing models and the development of the interactivist approach. Notably, this paper presents a different view on the development of the reflective abilities demonstrated by children passing the FBT and other ToM tests than the one presented in CM, which is a more mature perspective and one that I currently endorse. AvC follows a more Nelsonian view, according to which language provides the central mechanism whereby mentalizing develops; CM offers a more nuanced view, arguing that physiological developments that enable reflection in the child around age 4 are also necessary.

The third paper, "**Conventional minds: An interactivist perspective on social cognition and its enculturation**" (CM), builds on the groundwork laid by AvC and presents the most comprehensive articulation of the positive model proposed in the dissertation. It elaborates on how interactivism provides a dynamic, action-based framework for understanding the enculturation of social cognition. This paper focuses on the role of conventions, cultural practices, and social interaction in the development of socio-cognitive abilities. CM also refines and modifies the interactivist interpretation introduced in AvC, as mentioned in the previous paragraph, offering a fully developed account of the proposed model.

The fourth paper, "**Krytyka natywizmu jawnego i ukrytego w badaniach nad dziecięcymi teoriami umysłu**" (KN), tackles the issue of innateness, a topic only briefly touched upon in the other texts. This paper critiques both explicit and implicit nativist assumptions in ToM research, emphasizing the conceptual and theoretical problems associated with nativist explanations. Inspired by Susan Oyama's developmental systems theory and Stephen Jay Gould's critique of preformationism, KN argues for a view of development that is contingent, dynamic, and shaped by interaction and environmental factors, challenging the idea of preformed, innate cognitive structures.

The fifth paper, "**Beyond the mirror: An action-based model of knowing through reflection**" (BM), elaborates on the interactivist model of reflection. Among others, it discusses how reflective processes are central to the development of socio-cognitive abilities and how they relate to language, elaborating on the already mentioned issue of the relationship between language and reflection. The paper situates reflection as a crucial mechanism in the development of explicit social understanding, such as the ability to attribute mental states to

others. It discusses how socio-cultural factors interplay with reflective ability in the development of explicit mentalizing across different cultures.

Finally, the two short commentaries, "**Encodingism is not just a bad metaphor**" (EM) and "**Encultured minds, not error reduction minds**" (ERM), address specific theoretical issues related to the broader critique of encodingism. They are commentaries to Brette (2019) and Veissière et al. (2019) respectively. EM critiques the persistent reliance on encodingist metaphors in neuroscience, arguing that such frameworks are fundamentally flawed and misrepresent the nature of cognition. ERM, on the other hand, critiques the application of the Free-Energy Principle to social cognition, pointing out how its assumptions fail to capture the dynamic and enculturated nature of human minds. Together, these commentaries extend the foundational critique of encodingism to specialized contexts, reinforcing the theoretical commitments of the dissertation.

Throughout the articles, I make use of various methods depending on the nature of the particular task. Much of the work concerns reviewing and critically assessing the cross-culture data on ToM development and interpretative positions assumed by the traditional ToM frameworks. The foundational critique advanced against those frameworks follows Bickhard's more general critique of encodingism and consists in demonstrating how the implicit ontological presuppositions of the traditional paradigm force conceptual constraints onto what is conceivable within it, and those constraints are argued to be fatal to the task of making sense of enculturated development of social cognition. The presentation of the positive account developed in the dissertation involves a summary of its basics, contextualization in relation to other theories, explication of some of the elements of those other theories within the interactivist framework, interpretation of the existing empirical data from the model's perspective, and illustrative comparison with other similar models.

4. Supporting research and publications

Apart from the research project of this dissertation, I have pursued a number of tangential projects, dealing both with social cognition and the interactivist model. Together with Joanna Teske, I have published an article addressing the post-modernist radical skepticism characterizing contemporary literature studies, where we suggest that the interactivist model of cognition and social ontology offer a solution to the theoretical problems that motivate that skepticism (Teske & Mirski, 2023). This work falls within the ambit of literature studies and

its philosophy and discusses some aspects of the interactivist model that were only briefly touched upon in my dissertation.

Most centrally, however, I have been working on extending the account presented in this dissertation to the problem of emotion and its role in socio-cognitive development. This has resulted in the publication of one article and preparation of two that are awaiting publication. The published article concerns the use of Philosophy for Children (P4C) to improve children's understanding of emotions (Janczura, Jakub, Karczmarczyk, Gut, & Mirski, 2022). The first of the unpublished two offers a criticism parallel to the one found in this dissertation, but applied to appraisal theory of emotion causation – I argue that appraisal theory is rooted in an encodingist ontology of the mind, which leads to some of the same problems discussed in this dissertation regarding ToM models. The second paper applies the interactivist model of emotion to the issue of emotional modulation of socio-cognitive development, exploring such issues as emotion regulation and its relation to children's social understanding. The ultimate aim of this effort is to further develop the account presented in this dissertation and expand it into a broader research project that incorporates empirical testing at the intersection of emotion studies and theory of mind.

5. Summary of the main argument and results

The present section describes the central argument of my dissertation. I begin with a brief presentation of the meta-theoretical perspective adopted in my research, followed by an overview of the research area of Theory of Mind. After that, I discuss the main critical points of my dissertation, and finally I present the interactivist account, together with its implications for empirical research in social cognition and its development.

It has been accepted in philosophy of science that some metaphysics is *necessarily* held by every researcher, either explicitly or implicitly, and forms the basis for their research activity. Science works by trying to find out facts about its object, but this can be done only via hypotheses that are generated on the basis of what properties the studied reality is assumed or presupposed to have. If this is granted, then it is clear that it should matter a great deal what one's metaphysics is, as it contributes to the formulation of research questions, study design (e.g. what control conditions can be conceivably relevant), and drawing of conclusions.

One prominent position has been that the metaphysical hard core of research programs cannot be successfully examined and assessed, and that the issue of progressiveness of research

programs is more tractable: The best we can do is to follow a research program as long as it is capable of continuous growth (generating new theories of increasingly wider empirical scope that are legitimately linked together with a common heuristic), and only when it is clearly degenerating (i.e. fails to increase its empirical content in a continuous manner) should we abandon it (Lakatos, 1970). Another, contrasting view has been that ontological assumptions of a research program can in fact be examined and assessed in principle, despite the impossibility of their direct empirical testing (see Campbell & Bickhard, 1986, pp. 10–33). The present dissertation shares that latter view.

In this project, I assume that it is not only possible but advisable to look at how any given theory conceives of its object, whether it is conceptually coherent, whether it makes sense within the wider context of our knowledge, and whether its consequences for empirical pursuits are not misguided. Accordingly, my criticism against the traditional ToM accounts of social cognition and its enculturation is leveled against their metaphysical "hard core", rather than their empirical adequacy. This is argued to be a more powerful approach, since empirical adequacy is prone to formulation of auxiliary hypotheses, and indeed such responses to criticism can be found in the ToM literature (Helming, Strickland, & Jacob, 2014, 2016; Westra & Carruthers, 2017). So, despite the fact that ToM models can be modified to account for the cross-cultural data, my critique shows that the ontological commitments of the models lead them to quite absurd claims. Most importantly, it is argued that the ontology in question faces a theoretical impossibility when the issue of developmental emergence is concerned, which shines especially bright in the context of enculturation of the mind. In keeping with this foundational theme, the positive model I present as an alternative to the ToM models has a strong focus on ontology and how it overcomes the issues discussed in the criticism.

5.1. What is Theory of Mind?

One way of characterizing Theory of Mind is as an empirical research program that continues the tradition established by the first studies that introduced the construct. In the study that served as a germ for later ToM research, Premack and Woodruff (1978) tested chimpanzees on a task that required them to understand the intention behind an agent's action. The authors hypothesized that the ability tested by the task was a theory of mind: a mental mechanism allowing one to infer the mental states of another agent and use them to predict their actions. Since their primate participants passed the test, the authors concluded that the chimpanzee indeed had a theory of mind. A few years later, the issue of theory of mind was taken up in

developmental psychology: considering the philosophical criticism that Premack and Woodruff's paper received (see Fenici, 2017, pp. 143–144), Wimmer and Perner (1983) devised a task that required their child participants to understand the difference between true and false beliefs – the false-belief task (FBT). The original study was in the form of a story told to the child, later it was administered in the form of a puppet show (Baron-Cohen, Leslie, & Frith, 1985), and finally as a scene acted out by human actors (Leslie & Frith, 1988).[1] The results were reliably robust in all of these implementations and showed that children start passing the FBT only around age 4. Since then, many other paradigms – such as unexpected-contents (Gopnik & Astington, 1988) or unexpected-identity (Perner, Leekam, & Wimmer, 1987) – were devised. Generally, the different paradigms give comparable results (Liu, Wellman, Tardif, & Sabbagh, 2008; Wellman, Cross, & Watson, 2001), but there are some modifications to the task that make it easier or more difficult for children, resulting in different ages of passing (Gut, Haman, Gorbaniuk, & Chylińskia, 2020; Lewis & Osborne, 1990; Mitchell & Lacohée, 1991; Rubio-Fernández & Geurts, 2013; Sullivan & Winner, 1993).

Additionally, FBT inspired a wide spate of related tests and research strands that are usually aimed to address some unanswered question about the ToM mechanism. These include but are not limited to research on perspective taking (Catala, Mang, Wallis, & Huber, 2017; Edwards & Low, 2019; Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010), comparative research with non-human primates (Borg, 2018; Call & Tomasello, 2008; Povinelli & Vonk, 2004), children's pretend play (Lillard, 2001), or emotion comprehension (Pons, Harris, & Rosnay, 2004). An important attempt to go beyond the single concept of belief has been made by Wellman and colleagues and their ToM Scale methodology (originally presented in Wellman & Liu, 2004), which is discussed in texts WT and AvC (Liu et al., 2008).

Thinking about children's cognitive competences in terms of theories was common practice around the time the first ToM studies were conducted. Starting from the 1960s, developmental psychologists were collecting a growing amount of evidence of infants' cognitive competence, starting with their perception of objects and understanding of physics, through their understanding of animacy, and finally other minds (this marks the beginning of ToM research; Wimmer and Perner (1983)). This evidence of early competence clashed with

[1] Here is a short description of the original change-of-location study as found in Baron-Cohen et al. (1985): Sally and Ann are in the same room. Sally puts the ball in her basket and leaves the room. In her absence, Ann moves the ball into her own box. Sally comes back into the room and the key question asked is "where will Sally look for her ball?" Most children aged 4 are able to solve this task correctly while 3-year-olds generally fail.

the then extant views on infants that saw them as completely helpless and inept. This required an explanation. The inspiration to conceive of the paradoxical abilities as theories seems to have come from current philosophy of science as well as budding cognitive science (cf. Nelson, 2007). "Theory-ladenness" of all observation was an important claim in current philosophy of science (Suppe, 1977, pp. 125–232): Combating the neopositivistic views that assumed a theory-neutral observation language, world-view philosophers pointed out that even the simplest acts of observation are influenced by our prior knowledge – what we observe is always "theory-laden". Viewed from the perspective of these discussions, competent infants indeed seemed to have some theory that "laded" their behavior in experiments. This theory-based perspective chimed well with contemporary trends in cognitive science, exemplified most clearly by such models as Chomsky (1965), Fodor (1975), and Newell (1980), which viewed the mind as a computer comprised of sub-systems or mechanisms that manipulate mental representations (concepts) according to prescribed rules of inference (see Erdin, 2020; Gardner, 1985). These mechanisms essentially mirror the inferential structure of theories (understood in a neopositivist sense – as logical calculi) and so the idea fit neatly with the model of children theoreticians or "little scientists". Theory of mind has been proposed alongside theories of other domains of children's seemingly precocious competence, such as biology (Carey, 1985) and physics (Spelke, 1988).

Overall, the original FBT studies set the course for the decades of research to come, resulting in an overwhelming domination of the ToM perspective in research on social cognition today. Most centrally for us, the recent decades of that research have brought in a heightened interest in cross-cultural questions.

5.2. Socio-cultural factors in ToM development

Vinden (1996) was the first one to demonstrate that children pass the FBT (and thus are assumed to possess the concept of belief) at different ages depending on culture. She demonstrated that Junin Quechua children of ages 4-8 performed at chance on a FBT (this contradicted the earlier study of Avis and Harris (1991) with the Baka from Cameroon, which replicated Western results of age-4 passing). Later, Vinden (1999) investigated FBT performance in three non-Western cultures: (1) the Mofu of Cameroon, (2) the Tolai and (3) Tainae of Papua New Guinea. The results showed a clear lag in FBT; in contrast to the age of four in the West, children from the three non-Western groups passed it approximately around 7 (though there was additional variety between the groups). The number of studies

demonstrating a cross-cultural variance in FBT performance multiplied around that time, and now there are substantial data available from a number of different cultures. Children shift to above chance performance in the FBT around 6 years in Japan (Naito, 2014). In Pacific cultures, such as Samoa (Mayer & Träuble, 2012) and Vanuatu (Dixson, Komugabe-Dixson, Dixson, & Low, 2017), the age is roughly 8 years (cf. earlier results that suggested lack of the difference in Callaghan et al. (2005)). Gracia, Peterson, and Rosnay (2016) tested Filipino children aged 3-6 and discovered that only 15% of the oldest of the group passed, leaving the question about the age of passing in that culture open. Similarly, Nawaz, Hanif, and Lewis (2014) demonstrated a lag for Pakistani children compared to Western samples where 4-year-olds still performed at chance. Further, while Chinese children from mainland China perform similarly to their Western counterparts, children from Hong Kong demonstrate a 2-years lag, starting to pass FBT around 6 (see a large meta-analysis in Liu et al., 2008). There are also differences among Western cultures: Lecce and Hughes (2010) found much worse performance on FBT among Italian children aged 5-6 than among their British peers.

The cross-cultural variance in ToM tests has been hypothesized to be proximally the result of more fine-grained socio-cultural factors, which also have been studied extensively over the last decades. Taken together it seems that children perform better on ToM tasks when they have been exposed to social experiences that involve consideration of minds and their states. As is evident in the data, this exposure can be had via more mentalistic language the surrounds the child (Milligan et al., 2007), greater number of people whose needs and thoughts are contrasted (siblings), or more "mindful" parents that explain mental life to them (parenting strategies) (Devine & Hughes, 2019; Tompkins et al., 2018).

5.3. The traditional models of ToM

As already mentioned, ToM research emerged in the conceptual context where it was common practice to conceive of children's cognitive abilities as theories, or theory-like mechanisms. The three dominant ToM models (or theory theory models) are clearly part of this tradition, differing in the details rather than the overarching ontology of the mind. They are modular nativism, rational constructivism, and two-systems theory. I present them most fully in WT and AvC, but also in KN.

The three models offer contrasting interpretations of the cross-cultural variance in ToM test performance. Rational constructivists have held that the results support their model, arguing that since it develops differentially depending on socio-cultural context, ToM is an

ontologically constructed, domain-general theory. In fact, most of the cross-cultural studies have been conducted by the rational constructivist camp. The nativists, however, have countered these claims with a competence-performance argument: the socio-cultural variance of ToM test performance is a product of executive limitations rather than lack of understanding (Helming et al., 2014, 2016; Westra & Carruthers, 2017). Children are said to possess an innate ToM understanding, including understanding of false beliefs, but other, "performative" factors make it difficult to let that competence shine. Finally, the middle ground argued by the two-systems theory has been that it is their system 2 that is subject to socio-cultural influence, but system 1 develops and works independent of it.

5.4. The foundational criticism

The criticism extended in this dissertation pertains to a level of analysis lower than the debate that has been taking place between the above discussed theories. That is, I argue that their differences notwithstanding, all three models (and potentially others that are based on encodingism) commit to an ontology of mind that is fatally flawed to begin with, rendering their interpretations untenable not on an empirical basis, but rather theoretical. The criticism draws extensively on Mark Bickhard's general critique of the ontology in question, pointing out how it applies to the particular debate over ToM and culture as well. Different angles to the critique can be found in all the papers constituting this dissertation; below I present a short summary.

Encodingism. The core of Bickhard's argument is that the ontology in question – encodingism – is an intrinsically untenable model of natural cognition; due to the limitations inherent in their foundational assumptions about the nature of the mind, encodingist models lead to absurd claims, which are largely ignored in the bulk of cognitive sciences. Versions of the criticism can be found in other scholars' work (Gibson, 1983[1966]; Hutto & Myin, 2013, 2017; Piaget, 1971; Thelen & Smith, 2002/1996), but I believe that Bickhard's is the most precise and exhaustive, historically one of the earliest (Bickhard, 1980b, 1980a), and most importantly, a comprehensive positive framework is offered in the place of the criticized one.

Most fundamentally, encodingism assumes or presupposes that organisms are able to know the world due to correlational relationships between their internal states and the states of the outside world. In modern days, this is most often framed in terms of informational relationships, where "information" is understood in information-theoretic sense (Shannon & Weaver, 1964). This foundational assumption is argued to be fatally flawed as correlations are

ubiquitous in the world and by themselves do not constitute knowledge. It is only when there is a knowing agent that can learn about the correlational relationship between the two that we get a form of knowledge. On fear of infinite regress, the agent cannot represent the world via correlational relationships – a different form of knowing is necessary to lie at the base of natural cognition. There are many angles to the problem and many implications for theory building and empirical research, some of which I discuss in the papers of this dissertation, while multiple more are discussed in Bickhard's work (see, especially, Bickhard & Terveen, 1995).

Foundationalism in development. For my purposes, the most central of the problems begot by encodingism is *foundationalism*. Foundationalism is the view that cognitive development can only progress from a foundation of already contentful representations; no emergence out of non-psychological phenomena can be modeled within a foundationalist framework. This follows from encodingism since in order to establish first representations (i.e. first internal states that correlate with some element of the outside world and which are understood as such by the organism), the organism would have to be able to step outside of itself and check what its internal states correlate. This is clearly impossible, so a more acceptable option is to claim that there are some inborn or otherwise already given mental contents, which can be used as the starting point for development. When development is viewed in terms of theory construction, such starting representations are usually assumed to be concepts that are used to generate hypotheses and test them against the world, leading to the construction of more complex representations. Note, however, that this view necessitates that all future knowledge draws on the starting mental contents and remains a mere reconfiguration of them. What is impossible within a foundationalist framework is to offer a model of how mental content emerges out of non-contentful phenomena – *the foundation is a theoretical necessity*.

Allen and Bickhard (2013) have demonstrated that despite being competitors, nativist and empiricist views in developmental psychology are equally committed to foundationalism in virtue of their joint adherence to encodingism. The difference and main point of contention between the two camps is the richness of the assumed foundations, not their existence. Allen and Bickhard's criticism surveyed a wide range of research on development, demonstrating its ubiquity in the discipline. My own critique is more focused and pertains only to the field of theory of mind, which is not discussed by them. In texts AvC, KN, and CM, I have argued that the squabbles about the richness of the foundations notwithstanding, all three of the major ToM models are in fact committed to foundationalism due to their assumption of encodingist

ontology. In order to overcome that limitation, a non-encodingist account needs to be developed, which I offer in the present dissertation.

Nativism. Additionally, one seeming way out of foundationalism is addressed at length: nativism. As I argue in AvC and KN, the claim of innateness of the developmental foundations is explicitly made by modular nativists, while the two other accounts generally avoid talking about the origins of their initial concepts. However, they fail to offer an alternative origin story of the starting concepts, and some notable proponents of the model have actually embraced a more minimalist nativism (Gopnik, 2003). I argue that this is so due to the conceptual limitations of their adopted ontology. Nativism forms the main subject of KN, but I also discuss it in AvC. In there, I try to demonstrate that none of the meanings of the term deployed in the literature offers a tenable solution to foundationalism – whether in ontogeny or phylogeny, the conceptual impossibility for the emergence of the first concepts remains .

5.5. Implications of foundationalism for socio-cognitive development

As I argue in texts AvC and CM, foundationalism of the ToM mechanism accounts leads to some peculiar consequences as far as culture and its impact on development are concerned. I believe that this issue in fact brings out the flaws of the traditional accounts in light fuller than when development is considered in more general terms.

Culture as a modulator of internally specified development. First, culture necessarily boils down to a mere modulator of development in a space of potentiality determined by the initial concepts. Cultural variation in ToM development can be only in terms of different trajectories that follow from the starting foundations. This means that no genuinely cultural content can emerge when a ToM account is adopted, rendering the diversity of concepts that are found across different cultures and their folk psychologies mere blends of the foundational contents.

The uncultured nature of the ToM mechanism. Second, the assumption behind theory theory is that social cognition is most fundamentally made possible by the child's ability to attribute mental states to others. This means that culture can have no impact on the nature of that mechanism – it can only influence what concepts are constructed from the initial foundation and attributed to others, not the principles of the operation of the mechanism. The upshot is that even cultures that tend not to engage in mental attribution when they folk psychologically reason about others are assumed to still conduct mental state attributions and reason about others' behavior on their basis, though "implicitly". As a consequence, the incredible

coincidence seems to be that the culturally universal mechanism behind social cognition at large has the same form as Western folk psychology. My claim here is naturally that there is no coincidence here as this assumption is a necessary theoretical constraint of the framework.

Instrumental treatment of language. Third, language, a part of culture, in its dazzling diversity across the globe, can play only a rather limited role in socio-cognitive development within the theory theory models. The formation of the theory of mind mechanism is essentially an internal process; it is independent of language, though language can provide information useful to its development. As such, language within the ToM paradigm is treated instrumentally – it does not impact the form of the cognitive mechanism in question in any other way than providing content for it. This contrast starkly with the constitutive role that is ascribed to language in my own account.

Empirical implications of encodingism. Fourth, there are empirical implications that follow from the encodingist models. Socio-cultural factors, such as language or private epistemologies of caregivers are considered in empirical ToM research only inasmuch as they can be a source of information for the "child's theoretician"; that is, instrumentally. This tends to ignore the plethora of other ways in which culture can matter for the child's socio-cognitive development. For instance, the constitutive role of culture in the creation of the social reality cannot be discussed from the perspective of the ToM approaches – the focus remains solely on the presumably universal concepts of belief and desire, while other, culturally specific ways of thinking about others are most often ignored or interpreted as noise in empirical studies.

5.6. The interactivist model

The interactivist model has been created and developed over the years by a single man – Mark Bickhard. Historically, the model is in a direct line a descendant of American pragmatism and its action-based theory of truth. What Bickhard has essentially achieved is a naturalization of the American pragmatist notion of truth – the interactivist model is a model of cognition where knowing is modelled on acting in the world, rather than having a picture of it. Also, drawing inspiration from Piaget (esp. his later, action-based theorizing), Bickhard has made development a major theme in his theory. On that front, he has managed to show that his action-based model makes genuine emergence of mental content possible. Naturally, content is understood here in a new way, as the encodingist model of content is argued to be inherently untenable.

Normativity. The model's basics are presented most exhaustingly in CM, but also in AvC. Interactivism offers a model of ontological emergence of mental phenomena out of life, and of life out of inanimate processes (normative emergence). An exhaustive treatment of the metaphysics can be found in the works of the late Richard Campbell (R. J. Campbell, 1992, 2015). Most fundamentally, Bickhard has proposed that the ontological domain that exhibits normativity lies within conditions that are far from thermodynamic equilibrium (FFE). FFE processes need constant flow of energy in order to persist, and so the provision of that energy flow is seen as the most fundamental criterion for judging normativity, making for the most fundamental selection pressure in the emergence of the normative realm. What is special about life is that it is recursively self-maintaining – it is a FFE process that is not only organized in such a manner that ensures its continuous existence (self-maintenant), but also that can change its organization to remain self-maintenant across different environments (recursive self-maintenance: self-maintenance of the ability to self-maintain). With evolutionary time, living organisms have developed a dedicated organization for keeping track of what type of activity is possible in the situation the organism is currently in. This is where mental representation is found.

Mental representation. Mental representation follows the same FFE dynamics as life in general, but its function for the organism is different; its contribution is not directly to self-maintenance, but rather to recursive self-maintenance – it makes the organism's ability to identify the nature of its environment and adopt a fitting mode of functioning much more effective. The central concept required to understand how interactivist representing works is *implicit presupposition*. Traditionally, implicit presupposition has been associated with philosophers of language such as Frege or Russell, who discussed implicit presuppositions of sentences such as "the king of France is bald", which presupposes the king of France to actually exist. Implicit presupposition in this sense is the content that is not explicitly present in the meaning of the sentence, but nevertheless logically necessary for the sentence to make sense – it is presupposed by what is said, rather than openly said. Bickhard's implicit presupposition is similar but a more general concept – it applies not only to sentences but to any normative process (that is to FFE processes). And so life, for instance, presupposes the boundary conditions where it can thrive (e.g. availability of oxygen for animals), the action of grabbing a mug presupposes its weight and solidity, and so on. If what is implicitly presupposed by a normative process is not present, the process destabilizes and eventually collapses. Mental

representation began when the organisms started using implicit presuppositions of its interactions with the world for informational purposes – a visual scan that anticipates certain light structure has some of the same presuppositions of drinking from a mug, and so can be used to ensure that drinking is in fact possible. Further evolution led to the formation and finer differentiation of systems dedicated to such representing. In brief, then, Bickhard talks about mental representing in terms of an organization of anticipatory processes, organized in virtue of the relationships between their implicit presuppositions.

Development. Psychological development occurs via a variation and selective retention processes whereby the organism attempts different combinations of interactions with the world, establishes what works and how different interactions are related presuppositionally. Successful organizations (i.e. such that flow in an anticipated manner) are retained and become part of the organism's mind, forming the context for further learning. That is, the model is that of *evolutionary epistemology* (D. T. Campbell, 1974) and *recursive constructivism*.

Social ontology in interactivism. In his *Cognition, convention, and communication* (1980a), Bickhard presented his model of social ontology and communication (they entail one another). The basic idea is that social ontology is constituted by social conventions, and social conventions are understood as anticipatory agreement between agents. In order to anticipate each other's behavior, an agent interacting with another needs to take into consideration not only their representation of the situation, but also their representation of the agent's own representation of the situation, and vice versa, seemingly leading to a vicious circle of mutual characterizations. This is the famous coordination problem (Schelling, 1997[1963]). Bickhard, drawing on Brandom (2000/1994), argues that this problem is solved via joint variation and selection learning processes whereby coordination precedence leads to complementary future anticipations – establishment of a convention. Conventions build up as a group of people continue to interact with one another and the anticipatory agreement becomes part of their psychological *and sociological* organization, new members of the group are introduced to the already extant conventions, continuing their existence. Such established conventions constitute social ontology – an emergent, interpersonal realm of normativity that each member of the given culture co-constitutes, but which exhibits high levels of independence from any one person (cf. Berger & Luckmann, 1967).

Language in interactivism. Language is a special type of convention – it is a meta-convention, a convention for interacting with conventions. With language, one can manipulate

situation conventions – for instance, by uttering the simple "no!", one can effect a stop to whatever situation convention is occurring. This interactivist model of language belongs to what Bickhard (1980) terms "transformation" models of language, contrasting them with "picture models", which hold that language works by encoding meanings in utterances and sending them back and forth between interlocutors. It is the latter type of model that has traditionally dominated in linguistics. However, models of language similar to the interactivist one can be found in linguistics today (Kempson, Cann, Gregoromichelaki, & Chatzikyriakidis, 2016, 2017).

Reflection in interactivism. The interactivist model of reflection is one of the main aspects that make interactivism stand out among other "4E" models of cognition, which tend to struggle to account for internal thought. Reflection makes for the main subject of BM, but is briefly discussed in CM and AvC as well. As I have already mentioned, interactivist mental content is constituted as implicit presuppositions of processes that anticipate changes in sensorimotor patterns; that is, such processes that interact with the world. In order to represent what is presupposed about the world by such processes, a second interactive system is argued to be needed, such that will interact with the anticipatory organization itself and form anticipations about it. This allows the organism to not only represent the world, but also its own representation of the world. I argue that reflection makes possible for the child to reflect over the cultural ways of being and to begin forming his or hers folk psychological understanding.

5.7. The interactivist account of social cognition and ToM

In this section I go over the interactivist framing of the development of social cognition and folk psychology that I present in CM.

Infant theory of mind? Contrary to ToM accounts, especially the nativist ones, the social competences of infants are not due to their ability to represent and attribute other people's mental states. Rather, basic socio-cognitive abilities are due to growing conventionalization of the child's mind. Within the interactivist framework, it is possible to behave in a way that "honors" or *implicitly presupposes* other people's mental states without actually representing them, and infants are argued to do just that. Explicit theorizing about other minds, part of folk psychology, comes later, after the child has developed an ability to reflect (around year 4).

Cultural constitution of the social world. Culture – social ontology – is inseparably part of the person within interactivism, and so it is part of their socio-cognitive abilities. This forms a stark contrast to the consideration of culture in theory theory, where it is more of an

afterthought. The child's initial understanding of others is by forming joint situation conventions that are mutually intelligible. Early, this takes the form of such routines as changing the diaper, where children adjust their bodies to facilitate the process.[2] But with time children's growing mastery over the recurring types of situation conventions grows and starts to organize itself also around the presuppositions about other people's mental states.

False-belief understanding. As has already been stated, contrary to the ToM accounts, the interactivist model makes it possible for the child to act in accordance with others' false beliefs without representing them as such. Explicit understanding of false representing becomes possible when reflection is available to the child. However, reflection merely enables the construction of such explicit understanding – it does not cause it. Depending on the cultural context, the child's reflective construction of folk psychological understanding can follow different patterns, not necessarily focusing on the mastery of the concepts of belief and desire. This cultural variation in what aspects of social and mental life are foregrounded in folk psychology is argued to be responsible for why children from various cultures have problems with ToM tests such as the false-belief test – explicit understanding of that kind is simply not that important in those cultures.

Folk psychology. Folk psychology is argued to be a reflective convention – it is a culturally evolved way of reflecting over one's social knowledge that helps members of a given culture to coordinate activity. As such, it is a highly important construct, but it does not form the basis for social cognition as has been claimed in the ToM accounts.

Language and socio-cognitive development. Language is not (solely or mainly) a source of additional information about other's mental states. Rather, language is constitutive of much of social cognition. Children cognize others in terms of conventions and linguistic interactions available in them, and those have implicit presuppositions about various aspects of the world, including the mental states of the people involved. Theory of mind, understood as explicit understanding of other minds, part of folk psychology, is formed with the use of reflection on the basis of such implicit, language-constituted knowledge. That is, it is the exact reverse of the traditional approach, where the ToM mechanism is a prerequisite for linguistic knowledge and interaction.

[2] Apparently, there are differences across different families, different styles of body adjusting etc.

Culture and socio-cognitive development. The view on culture and development that follows from interactivism is that children first develop interactive knowledge about the world that implicitly presupposes various properties of that world without representing them explicitly. Social knowledge of this kind is constituted mostly of situation conventions that make up the child's culture. The child thus grows into the culture, becomes part of it, rather than being merely influenced by it. The socio-cultural knowledge implicitly presupposes multiple facts about other people and their minds, some of which are abstracted and reflected over when the child becomes capable of reflection. The existence of folk psychology in the culture helps this process as it foregrounds those aspects of social life that are of most importance in that culture. Belief-desire understanding is only a small part of social cognition, a part that is important to varying extent across different cultures, leading to differential performance on ToM tasks in them.

5.8. Conclusion

The main conclusions that follow from my doctoral project are that (1) Theory of Mind models are inherently flawed due to their ontological assumptions and cannot offer an adequate account of socio-cultural development, regardless of their consistency with empirical data. And (2), the interactivist ontology and the account developed in this dissertation form a better alternative to those models as they overcome the limitations in question.

Regarding the first point, the problematic ontological assumptions of the ToM models make psychological emergence impossible, forcing the models to postulate a set of foundational concepts. This, in turn, results in a reductive and untenable account of cultural influence on socio-cognitive development. While the criticism extended in the dissertation may strike one as overly assertive, it needs to be stressed that if the argument is taken seriously, it indeed forces such strong conclusions. This naturally does not mean that all the empirical results of the research programs based on the ToM models are invalid: As I demonstrate, they can be given other interpretations, such as the interactivist one offered in this dissertation, and the methodological problems that follow from the faulty ontology can be addressed in future research.

The interactivist account offered in place of the ToM models provides a dynamic, action-based framework that successfully integrates socio-cognitive development and cultural influence without relying on reductive foundational concepts. By emphasizing the emergent, interactive nature of cognition and its deep embedding in socio-cultural contexts, this approach

resolves the conceptual issues inherent in traditional ToM frameworks. Furthermore, it opens new avenues for interpreting existing empirical data and designing research that more accurately captures the complexity of socio-cognitive processes.

In summary, this dissertation not only critiques the limitations of dominant ToM models but also offers a robust alternative that aligns theoretical coherence with empirical applicability. While further work is needed to refine and expand the interactivist account, the foundations laid here provide a strong basis for rethinking the study of socio-cognitive development. The hope is that this work will encourage more integrative, interdisciplinary approaches that transcend the constraints of current paradigms, fostering a richer understanding of how culture and interaction shape the human mind.

6. References

Allen, J. W., & Bickhard, M. H. (2013). Stepping off the pendulum: Why only an action-based approach can transcend the nativist–empiricist debate. *Cognitive Development, 28*(2), 96–133. https://doi.org/10.1016/j.cogdev.2013.01.002

Avis, J., & Harris, P. L. (1991). Belief-Desire Reasoning among Baka Children: Evidence for a Universal Conception of Mind. *Child Development, 62*(3), 460–467. https://doi.org/10.1111/j.1467-8624.1991.tb01544.x

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind" ? *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Berger, P. L., & Luckmann, T. (1967). *The social construction of reality: A treatise in the sociology of knowledge* (Anchor books ed.). *Anchor books: A 589*. Garden City, NY: Doubleday.

Bickhard, M. H. (1980a). *Cognition, convention, and communication*. New York, NY: Praeger.

Bickhard, M. H. (1980b). A Model of Developmental and Psychological Processes. *Genetic Psychology Monographs*, 61–116.

Bickhard, M. H. (2009). The interactivist model. *Synthese, 166*(3), 547–591. https://doi.org/10.1007/s11229-008-9375-x

Bickhard, M. H. (2016). Inter- and En- activism: Some thoughts and comparisons. *New Ideas in Psychology*, *41*, 23–32. https://doi.org/10.1016/j.newideapsych.2015.12.002

Bickhard, M. H. (2024). *Whole Person: Contributions to an Ontological Psychology*. [S.l.]: ELSEVIER ACADEMIC PRESS.

Bickhard, M. H., & Richie, D. M. (1983). *On the nature of representation: A case study of James Gibson's theory of perception / Mark H. Bickhard, D. Michael Richie with Robert Hughes, James Dannemiller*. New York, NY, USA: Praeger. https://doi.org/10.1057/9781137400437.0005

Bickhard, M. H., & Terveen, L. (1995). *Foundational issues in artificial intelligence and cognitive science: Impasse and solution*. *Advances in psychology: Vol. 109*. Amsterdam: Elsevier.

Borg, E. (2018). On Deflationary Accounts of Human Action Understanding. *Review of Philosophy and Psychology*, *9*(3), 503–522. https://doi.org/10.1007/s13164-018-0386-3

Brandom, R. (2000/1994). *Making it explicit: Reasoning, representing, and discursive commitment* (3rd pr). Cambridge, Mass. [u.a.]: Harvard University Press.

Brette, R. (2019). Is coding a relevant metaphor for the brain? *Behavioral and Brain Sciences*, 1–44. https://doi.org/10.1017/S0140525X19000049

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, *12*(5), 187–192. https://doi.org/10.1016/j.tics.2008.02.010

Callaghan, T., Rochat, P., Lillard, A. S., Claux, M. L., Odden, H., Itakura, S., . . . Singh, S. (2005). Synchrony in the Onset of Mental-State Reasoning: Evidence From Five Cultures. *Psychological Science*, *16*(5), 378–384. https://doi.org/10.1111/j.0956-7976.2005.01544.x

Campbell, D. T. (1974). Evolutionary Epistemology. In P. A. Schilpp (Ed.), *The philosophy of Karl Popper* (pp. 413–463). La Salle, Illinois: Open Court.

Campbell, R. J. (1992). *Truth and historicity*. Oxford: Clarendon.

Campbell, R. J. (2015). *The metaphysics of emergence*. New York, NY: Palgrave Macmillan.

Carey, S. (1985). *Conceptual change in childhood*. *The MIT Press series in learning, development, and conceptual change*. Cambridge, Mass.: MIT Press.

Catala, A., Mang, B., Wallis, L., & Huber, L. (2017). Dogs demonstrate perspective taking based on geometrical gaze following in a Guesser-Knower task. *Animal Cognition.* Advance online publication. https://doi.org/10.1007/s10071-017-1082-x

Chomsky, N. (1965). *Aspects of the theory of syntax* (25. print). *Special technical report of the Research Laboratory of Electronics of the Massachusetts Institute of Technology: Vol. 11*. Cambridge, Mass.: MIT Press.

Devine, R. T., & Hughes, C. [Claire] (2019). Let's Talk: Parents' Mental Talk (Not Mind-Mindedness or Mindreading Capacity) Predicts Children's False Belief Understanding. *Child Development*, *90*(4), 1236–1253. https://doi.org/10.1111/cdev.12990

Dixson, H. G. W., Komugabe-Dixson, A. F., Dixson, B. J., & Low, J. (2017). Scaling Theory of Mind in a Small-Scale Society: A Case Study From Vanuatu. *Child Development.* Advance online publication. https://doi.org/10.1111/cdev.12919

Edwards, K., & Low, J. (2019). Level 2 perspective-taking distinguishes automatic and non-automatic belief-tracking. *Cognition*, *193*, 104017. https://doi.org/10.1016/j.cognition.2019.104017

Erdin, H. O. (2020). Appraisal of certain methodologies in cognitive science based on Lakatos's methodology of scientific research programmes. *Synthese.* Advance online publication. https://doi.org/10.1007/s11229-020-02612-4

Fenici, M. (2017). Rebuilding the Landscape of Psychological Understanding After the Mindreading War. Advance online publication. https://doi.org/10.13128/Phe_Mi-21113

Fodor, J. A. (1975). *The language of thought*. *The language and thought series*. New York, NY: Crowell.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution / Howard Gardner*. New York: Basic.

Gibson, J. J. (1983[1966]). *Senses considered as perceptual systems*. Westport, Conn: Greenwood.

Gopnik, A. (2003). The Theory Theory as an Alternative to the Innateness Hypothesis. In L. M. Antony & N. Hornstein (Eds.), *Philosophers and their critics. Chomsky and his critics* (pp. 238–254). Malden, MA: Blackwell Pub. https://doi.org/10.1002/9780470690024.ch10

Gopnik, A., & Astington, J. W. (1988). Children's Understanding of Representational Change and Its Relation to the Understanding of False Belief and the Appearance-Reality Distinction. *Child Development*, *59*(1), 26–37. https://doi.org/10.1111/j.1467-8624.1988.tb03192.x

Gracia, M. R. L. de, Peterson, C. C., & Rosnay, M. de (2016). A Cultural Conundrum. *Journal of Cross-Cultural Psychology*, *47*(7), 929–940. https://doi.org/10.1177/0022022116655790

Gut, A., Haman, M., Gorbaniuk, O., & Chylińskia, M. (2020). The Development of Understanding Opacity in Preschoolers: A Transition From a Coarse- to Fine-Grained Understanding of Beliefs. *Frontiers in Psychology*, *11*, 596. https://doi.org/10.3389/fpsyg.2020.00596

Helming, K. A., Strickland, B., & Jacob, P. (2014). Making sense of early false-belief understanding. *Trends in Cognitive Sciences*, *18*(4), 167–170. https://doi.org/10.1016/j.tics.2014.01.005

Helming, K. A., Strickland, B., & Jacob, P. (2016). Solving the Puzzle about Early Belief-Ascription. *Mind & Language*, *31*(4), 438–469. https://doi.org/10.1111/mila.12114

Hutto, D. D., & Myin, E. (2013). *Radicalizing enactivism: Basic minds without content / Daniel D. Hutto and Erik Myin*. Cambridge, MA: MIT Press.

Hutto, D. D., & Myin, E. (2017). *Evolving enactivism: Basic minds meet content*. Cambridge Massachusetts: MIT Press.

Janczura, Jakub, Karczmarczyk, A., Gut, A., & Mirski, R. (2022). W stronę lepszego rozumienia emocji: wpływ filozofowania z dziećmi na uczenie emocji – przegląd badań rozwojowych. *Roczniki Filozoficzne*, *LXX*(3), 109–139.

Kempson, R., Cann, R., Gregoromichelaki, E., & Chatzikyriakidis, S. (2016). Language as Mechanisms for Interaction. *Theoretical Linguistics*, *42*(3-4), 203–276. https://doi.org/10.1515/tl-2016-0011

Kempson, R., Cann, R., Gregoromichelaki, E., & Chatzikyriakidis, S. (2017). Action-Based Grammar. *Theoretical Linguistics*, *43*(1-2), 141–167. https://doi.org/10.1515/tl-2017-0012

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos (Ed.), *Criticism and the growth of knowledge* (pp. 91–195). S.l.: CUP.

Lavelle, J. S. (2019). The impact of culture on mindreading. *Synthese*, *87*(3), 1–24. https://doi.org/10.1007/s11229-019-02466-5

Lecce, S., & Hughes, C. [Claire.] (2010). The Italian job? Comparing theory of mind performance in British and Italian children. *British Journal of Developmental Psychology*, *28*(4), 747–766. https://doi.org/10.1348/026151009X479006

Leslie, A. M., & Frith, U. (1988). Autistic children's understanding of seeing, knowing and believing. *British Journal of Developmental Psychology*, *6*(4), 315–324. https://doi.org/10.1111/j.2044-835X.1988.tb01104.x

Lewis, C., & Osborne, A. (1990). Three-Year-Olds' Problems with False Belief: Conceptual Deficit or Linguistic Artifact? *Child Development*, *61*(5), 1514. https://doi.org/10.2307/1130760

Lillard, A. (2001). Pretend Play as Twin Earth: A Social-Cognitive Analysis. *Developmental Review*, *21*(4), 495–531. https://doi.org/10.1006/drev.2001.0532

Liu, D., Wellman, H. M., Tardif, T., & Sabbagh, M. A. (2008). Theory of mind development in Chinese children: A meta-analysis of false-belief understanding across cultures and languages. *Developmental Psychology*, *44*(2), 523–531. https://doi.org/10.1037/0012-1649.44.2.523

Mayer, A., & Träuble, B. (2012). Synchrony in the onset of mental state understanding across cultures? A study among children in Samoa. *International Journal of Behavioral Development*, *37*(1), 21–28. https://doi.org/10.1177/0165025412454030

Milligan, K., Astington, J. W., & La Dack (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, *78*(2), 622–646. https://doi.org/10.1111/j.1467-8624.2007.01018.x

Mitchell, P., & Lacohée, H. (1991). Children's early understanding of false belief. *Cognition*, *39*(2), 107–127. https://doi.org/10.1016/0010-0277(91)90040-B

Naito, M. (2014). From Theory of Mind to Theory fo Relation: Sociocultural Perspectives on Japanese Children's Social Understanding. In O. N. Saracho (Ed.), *Contemporary Perspectives in Early Childhood Education. Contemporary perspectives on research in theory of mind in early childhood education* (pp. 381–408). Charlotte, North Carolina: Information Age Publishing.

Nawaz, S., Hanif, R., & Lewis, C. (2014). 'Theory of mind' development of Pakistani children: Do preschoolers acquire an understanding of desire, pretence and belief in a universal sequence? *European Journal of Developmental Psychology*, *12*(2), 177–188. https://doi.org/10.1080/17405629.2014.973843

Nelson, K. (2007). *Young minds in social worlds: Experience, meaning, and memory / Katherine Nelson*. Cambridge, Mass., London: Harvard University Press.

Newell, A. (1980). Physical Symbol Systems *Cognitive Science: A Multidisciplinary Journal*, *4*(2), 135–183. https://doi.org/10.1207/s15516709cog0402_2

Perner, J. [Josef], Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, *5*(2), 125–137. https://doi.org/10.1111/j.2044-835X.1987.tb01048.x

Piaget, J. (1971). *Genetic epistemology*. New York: Norton.

Pons, F., Harris, P. L., & Rosnay, M. de (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology*, *1*(2), 127–152. https://doi.org/10.1080/17405620344000022

Povinelli, D. J., & Vonk, J. (2004). We Don't Need a Microscope to Explore the Chimpanzee's Mind. *Mind & Language*, *19*(1), 1–28. https://doi.org/10.1111/j.1468-0017.2004.00244.x

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(04), 515. https://doi.org/10.1017/S0140525X00076512

Rubio-Fernández, P., & Geurts, B. (2013). How to pass the false-belief task before your fourth birthday. *Psychological Science*, *24*(1), 27–33. https://doi.org/10.1177/0956797612447819

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology. Human Perception and Performance*, *36*(5), 1255–1266. https://doi.org/10.1037/a0018729

Schelling, T. C. (1997[1963]). *Strategy of conflict* (2. edition, 16. printing). London: Harvard University.

Shannon, C. E., & Weaver, W. (1964). *The mathematical theory of communication*. Urbana: University of Illinois Press.

Slaughter, V., & Perez-Zapata, D. (2014). Cultural Variations in the Development of Mind Reading. *Child Development Perspectives*, *8*(4), 237–241. https://doi.org/10.1111/cdep.12091

Spelke, E. S. (1988). The origins of physical knowledge. In L. Weiskrantz (Ed.), *A Fyssen Foundation Symposium. Thought without language: 3rd Symposium : Papers and discussions / edited by L. Weiskrantz*. Oxford: Clarendon.

Sullivan, K., & Winner, E. (1993). Three-year-olds' understanding of mental states: The influence of trickery. *Journal of Experimental Child Psychology*, *56*(2), 135–148. https://doi.org/10.1006/jecp.1993.1029

Suppe, F. (1977). *The structure of scientific theories* (2nd ed. / edited with a critical introduction and an afterword by Frederick Suppe). Urbana, London: University of Illinois Press.

Szpak, M., & Białecka-Pikul, M. (2020). Links between attachment and theory of mind in childhood: Meta-analytic review. *Social Development*, *29*(3), 653–673. https://doi.org/10.1111/sode.12432

Taumoepeau, M. (2019). Culture, Communication and Socio-cognitive Development: Understanding the Minds of Others. In T. Tulviste, D. L. Best, & J. L. Gibbons (Eds.), *Children's Social Worlds in Cultural Context* (Vol. 78, pp. 41–54). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-27033-9_4

Teske, J., & Mirski, R. (2023). Interactivist Refutation of Radical Epistemic Skepticism Characteristic of Postmodern Literary Studies and Fiction. *ENTHYMEMA*. (32), 1–27. https://doi.org/10.54103/2037-2426/19781

Thelen, E., & Smith, L. B. (2002/1996). *A dynamic systems approach to the development of cognition and action* (5. print). *A Bradford book*. Cambridge, Mass.: MIT Press.

Tompkins, V., Benigno, J. P., Kiger Lee, B., & Wright, B. M. (2018). The relation between parents' mental state talk and children's social understanding: A meta-analysis. *Social Development*, *78*, 1–24. https://doi.org/10.1111/sode.12280

Veissière, S. P. L., Constant, A., Ramstead, M. J. D., Friston, K. J., & Kirmayer, L. J. (2019). Thinking Through Other Minds: A Variational Approach to Cognition and Culture. *Behavioral and Brain Sciences*, 1–97. https://doi.org/10.1017/S0140525X19001213

Vinden, P. G. (1996). Junin Quechua Children's Understanding of Mind. *Child Development*, *67*(4), 1707–1716. https://doi.org/10.1111/j.1467-8624.1996.tb01822.x

Vinden, P. G. (1999). Children's Understanding of Mind and Emotion: A Multi-culture Study. *Cognition & Emotion*, *13*(1), 19–48. https://doi.org/10.1080/026999399379357

Wellman, H. M. (2018). Theory of mind: The state of the art. *European Journal of Developmental Psychology*, *7*, 1–28. https://doi.org/10.1080/17405629.2018.1435413

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief. *Child Development*, *72*(3), 655–684. https://doi.org/10.1111/1467-8624.00304

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Westra, E., & Carruthers, P. (2017). Pragmatic development explains the Theory-of-Mind Scale. *Cognition*, *158*, 165–176. https://doi.org/10.1016/j.cognition.2016.10.021

Wimmer, H., & Perner, J. [J.] (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

7. Articles

Below can be found the articles that compose the present dissertation. They are arranged with respect to their content as discussed in section 3.

Robert Mirski

Wydział Filozofii
Katolicki Uniwersytet Lubelski Jana Pawła II

WSPÓŁCZESNE TRUDNOŚCI I WYZWANIA W MYŚLENIU O TEORII UMYSŁU – PERSPEKTYWA MIĘDZYKULTUROWA[1]

1. Wstęp

Myśli, uczucia czy intencje skrywające się w umysłach innych ludzi są jednym z głównych tematów rozmyślań każdego człowieka. Społeczna natura ludzkiego życia sprawia, że zdolność do rozumienia przekonań, pragnień oraz celów innych ludzi jest kluczowa dla przetrwania oraz szczęścia nie tylko jednostki, ale i całego gatunku. Nie jest więc zaskakujące, że problem innych umysłów sięga daleko w przeszłość filozofii, do Augustyna, Milla, Schelera czy Ryle'a. Współcześnie najbardziej płodnym obszarem badań eksplorującym ludzką zdolność do poznania społecznego jest przedsięwzięcie naukowe w ramach badań kognitywnych (szczególnie połączenie badań filozoficznych i psychologicznych), które formułuje ten problem w kategoriach potocznej teorii umysłu mającej umożliwiać ludziom wgląd w psychikę innych przedstawicieli *homo sapiens*. Pomimo tego, że istnieje szereg alternatywnych teorii tłumaczących zdolności poznawcze stojące za ludzkim poznaniem społecznym, pogląd postulujący, że zasadza się ono na strukturze poznawczej o formie teorii – stanowisko zwane teorią teorii – dominuje w literaturze (Wellman 2014; Gut 2016).

Celem niniejszego artykułu jest przedstawienie problemów, na jakie napotyka teoria teorii w świetle badań etnograficznych i rozwojowych dotyczących poznania społecznego zebranych w kulturach odmiennych od Zachodu. Pogląd, że poznanie społeczne opiera się na uniwersalnej, implicytnej teorii umysłu, niezależnie od dominującej w danej kulturze psychologii potocznej, jest zagrożony przez wyniki badań międzykulturowych, które demonstrują różnice

[1] Niniejszy tekst został przygotowany w ramach projektu badawczego pt. „O społeczno-kulturowym uwikłaniu potocznych teorii umysłu. Dyskusja o granicach uniwersalności i powszechności pojęć i atrybucji mentalnych", finansowanego przez Narodowe Centrum Nauki (grant PRELUDIUM 12, UMO-2016/23/N/HS1/02887 na lata 2017-2019).

w zdolnościach poznania społecznego wśród członków kultur z innymi psychologiami potocznymi niż dominująca na Zachodzie teoria umysłu. Proponowane przez teoretyków teorii interpretacje tych wyników mające na celu zachować tezę o uniwersalizmie implicytnej teorii umysłu napotykają jednak na dalsze problemy, które przedstawiam w dalszej części artykułu. Na koniec sygnalizuję alternatywne ramy teoretyczne, które mogą okazać się efektywniejsze w wyjaśnianiu poznawczych zdolności, na których bazuje ludzkie poznanie społeczne, oraz ich rozwoju.

2. Terminologia

Przed przejściem do ekspozycji problemu dokonam krótkich ustaleń terminologicznych, niezbędnych dla prezentowanych kwestii. Pomimo faktu, że terminy „teoria umysłu" i „potoczna psychologia" są często używane zamiennie, ze względu na podejmowany tutaj problem dokonam terminologicznego rozróżnienia, które przedstawiam poniżej.

Potoczną psychologię będę definiował jako zdroworozsądkową i powszechną teorię wyjaśniającą ludzkie zachowania w kategoriach intencjonalnych (Churchland 1998), bez precyzowania jej zestawu pojęć i istniejących pomiędzy nimi relacji. Poprzez teorię umysłu będę rozumiał już konkretny przypadek potocznej psychologii – potoczną psychologię panującą w świecie Zachodu, teorię, która jest konstytuowana przez pojęcia takie jak umysł, przekonanie, pragnienie itp. Posiadający teorię umysłu podmiot interpretuje zachowanie innych podmiotów w kategoriach posiadanego przez nich umysłu, który skrywa przekonania i przyjmuje w stosunku do nich pewne postawy (Wellman 2014; Gopnik, Wellman 1992).

Ponadto w tekście często będę się odnosił do „testu fałszywych przekonań" (*false belief test*; Wimmer 1983). Wyniki badań przeprowadzonych przy użyciu tego testu są głównym źródłem dyskutowanej przeze mnie kontrowersji. Zdanie testu zwykło się przyjmować za warunek posiadania przez daną osobę teorii umysłu. Istnieje kilka wersji testu, ale we wszystkich naczelnym pytaniem jest to, czy badany rozumie, że jeśli ktoś posiada fałszywe przekonanie, to będzie zachowywał się zgodnie z nim, a nie z rzeczywistością. Ta główna zasada jest najlepiej widoczna w wersji, gdzie badany siedzi w pokoju przy stole, na którym jest pudełko, wiaderko oraz zabawka. W pokoju są jeszcze dwie osoby – Kasia i Tomek. W obecności Tomka, Kasia umiejscawia zabawkę w pudełku, po czym opuszcza ona pokój. Podczas nieobecności Kasi, Tomek przekłada zabawkę z pudełka do wiaderka. Następnie Kasia wraca i badanemu zadaje się pytanie, gdzie Kasia będzie szukać zabawki najpierw. Aby zdać, badany musi odpowiedzieć zgodnie z fałszywym przekonaniem Kasi i wskazać pudełko jako jej pierwszy cel. W niektórych badaniach dodatkowym kryterium zdania testu jest wyjaśnie-

nie swojego wyboru przy użyciu terminologii odnoszącej się do umysłu i stanów mentalnych.

3. Teoria teorii

Pomimo tego, że istnieje obecnie kilka alternatywnych ram teoretycznych opisujących zdolność do „czytania" innych umysłów, to dominującą, szczególnie w psychologii poznawczej i rozwojowej, jest teoria teorii[2]. Zakłada się w niej, że zdolność do rozumienia zachowania innych ludzi w kategoriach intencjonalnych jest możliwa dzięki posiadanej przez człowieka teorii umysłu, czyli zestawu pojęć i heurystyk, które aplikuje on do obserwowanej rzeczywistości, aby zrozumieć i przewidzieć zachowania innych ludzi. W ujęciu tym dane percepcyjnie jest jedynie samo zachowanie podmiotu (rozróżnienie podmiot – przedmiot jest obecne na poziomie percepcyjnym), toteż aby zrozumieć to zachowanie, potrzeba człowiekowi teorii postulującej istnienie nieobserwowalnych bytów wyjaśniających obserwowaną rzeczywistość – pragnień, przekonań itd.

Oczywistym jest, że potoczna psychologia z przyjętej wcześniej przeze mnie definicji musi być teorią, oraz to, że w świecie zachodnim jest ona niezaprzeczalnie teorią umysłu. Sama ta kwestia jest trywialna. Jednak teoretycy teorii formułują silniejszy postulat, że większość naszego poznania społecznego jest umożliwiona dzięki temu, że dysponujemy taką teorią umysłu; używamy jej nie tylko wtedy, kiedy otwarcie opisujemy czyjeś zachowanie lub staramy się dociec, dlaczego dana osoba zachowała się w pewien kuriozalny sposób, ale także wtedy, gdy widzimy przechodnia schylającego się po monetę leżącą na chodniku lub wymieniamy grzeczności z naszą sąsiadką. Innymi słowy, teoria umysłu jest niezbędna do jakiegokolwiek zrozumienia społecznego nie tylko wtedy, kiedy eksplicytnie dokonujemy inferencji i stawiamy hipotezy na temat nietypowego zachowania, ale również wtedy, kiedy podświadomie rozumiemy celowość zachowania drugiego człowieka. Teoria umysłu nie jest więc tylko nabytkiem kulturowym, wyuczonym sposobem, w jaki ludzie zwykli sobie tłumaczyć swoje wzajemne zachowanie, ale także pierwotniejszą strukturą poznawczą, która sprawia, że takie potoczne psychologizowanie jest w ogóle możliwe.

W efekcie otrzymujemy dwa znaczenia „teorii umysłu": teoria umysłu jako psychologia potoczna, która jest *ipso facto* teorią, oraz teoria umysłu jako rozwojowo pierwotna struktura poznawcza umożliwiająca poznanie społeczne,

[2] Drugim, równie powszechnym stanowiskiem jest teoria symulacji. W obecnym artykule jej jednak nie podejmuję za względu na klarowność ekspozycji problematyki oraz konkluzywną w moim mniemaniu krytykę symulacyjnego podejścia obecną w literaturze (Gallagher 2007, zob. też Newen 2015). Niemniej jednak część przedstawianych przeze mnie problemów można również zarzuć symulacjonizmowi.

czyli m.in. również nabycie eksplicytnej teorii psychologii potocznej. To, że ta struktura poznawcza jest w pewnym przynajmniej stopniu niezależna od języka i kultury, zostało już przyjęte w oryginalnym badaniu Premacka i Woodruffa (1978), gdzie autorzy pytali, czy szympansy posiadają teorię umysłu. Wyszli oni z założenia, że percepcja nie zawiera informacji o stanach mentalnych innych ludzi, a więc musi być wywnioskowana przy pomocy teorii. Szympansy wykazują pewien stopień zrozumienia psychiki człowieka, stąd nie może być taka niezbędna teoria wyuczona językowo.

Konsekwencją tego poglądu jest więc potrzeba wyjaśnienia, skąd ta teoria umysłu bierze się w umyśle człowieka (i innych zwierząt). Naturalnym w tej sytuacji ruchem (i w gruncie rzeczy jedynym, zob. Allen, Bickhard 2013) jest sięgnięcie do długiej tradycji filozofii umysłu i postulowanie jej wrodzoności (np. Fodor 1992, 1983, 1975). Kwestia tego, jak wiele z teorii umysłu jest wrodzone, a jak wiele jest nabyte podczas rozwoju, jest przedmiotem toczących się dyskusji (np. Westra, Carruthers 2017). Zasadnicze dla niniejszego artykułu jest jednak to, że niezależnie od pozycji na skali wrodzone – nabyte, zawsze postuluje się pewien zakres wiedzy – fundament reprezentacyjny – na którym zasadza się dalszy rozwój poznawczy[3]. Uniwersalność teorii umysłu jako implicytnej struktury poznawczej jest więc z pozoru możliwa nawet w obliczu braku uniwersalności teorii umysłu jako psychologii potocznej, który został stwierdzony i który omawiam w następnej sekcji.

4. Czy potoczna psychologia równa się teorii umysłu?

Teoria umysłu jako psychologia potoczna, czyli eksplicytny sposób wyjaśniania zachowań ludzkich, nie jest zjawiskiem uniwersalnym kulturowo. Obserwujemy szereg kultur, w których psychologia potoczna przybiera formę inną niż teoria umysłu. Poniżej przedstawiam najbardziej jaskrawe przykłady.

Kultura panująca na Samoa i innych wyspach Pacyfiku charakteryzuje się psychologią potoczną, którą Joel Robbins i Alan Rumsey określają jako „nie-

[3] Zdaję sobie sprawę, że pojęcia uniwersalności i wrodzoności nie są tożsame. Ponadto przydatność samego pojęcia wrodzoności jest wysoce wątpliwa (Oyama i wsp. 2000; Mameli i Bateson 2011). Jednakże w rozumieniu wrodzoności używanym w literaturze teorii teorii to, co wrodzone, zawsze jest uniwersalne (wrodzoność zawsze prowadzi do uniwersalności, natomiast uniwersalność niekoniecznie oznacza wrodzoność). Niezależnie od krytyki takiego stanowiska, teoretycy teorii postrzegają wrodzone struktury poznawcze jako rozwijające się niezależnie od kontaktu z drugim człowiekiem i kulturą, opierając swoje stanowisko na argumentach odnoszących się do procesów adaptacyjnych w filogenezie. Wrodzona teoria umysłu będzie więc obecna w psychice każdego człowieka „ze względu na jej wrodzoność niezależnie od kontekstu rozwoju". Podejmuję ten temat w dalszej części tekstu. Dziękuję anonimowemu recenzentowi za zwrócenie uwagi na potrzebę klaryfikacji tej kwestii.

przeniknioność innych umysłów" (*the opacity of other minds*, Robbins, Rumsey 2008). Na wyspach panuje przekonanie, że jest praktycznie niemożliwym poznać myśli drugiej osoby, a nawet swoje własne. Samoanie niechętnie mówią o swoich stanach mentalnych i tych doświadczanych przez innych ludzi, odmawiając na przykład wyjaśnienia swojego zachowania (Duranti 2008; Mayer, Trauble 2012, 22-23). W zdecydowanym przeciwieństwie do zachodniego świata, interakcje społeczne zachodzą na Pacyfiku z założeniem, że niemożliwym jest odgadnięcie, co druga osoba myśli. Doktryna o nieprzeniknioności umysłów jest tam szeroko rozpowszechnionym, przyjmowanym za pewnik poglądem, który przenika całą kulturę, nadając kształt codziennemu życiu (Robbins, Rumsey 2008, 411).

Kolejnym przykładem kultury z odmienną od Zachodu psychologią potoczną jest Japonia. Mika Naito (2014) argumentuje za potrzebą kulturowo-specyficznego interpretowania czytania w umyśle, przedstawiając sposób, w jaki teoria życia mentalnego konstruowana w kulturze Japonii jest raczej teorią relacji, która dotyczy ogólnej intersubiektywnej „atmosfery" (*kokoro*), a nie poszczególnych umysłów, tak jak zwykło się ją opisywać w badaniach zdominowanych przez Zachód. Japończycy zwykli w swoich wyjaśnieniach zachowania odnosić się do norm społecznym i międzyludzkich relacji, nie do stanów mentalnych przypisywanych podmiotom.

Vinden (1996) pokazuje, że lud Keczua z regionu Junin w Peru w ogóle nie posiada psychologii potocznej jako takiej. W sytuacjach, w których mieszkaniec Zachodu zastanawiałby się nad tym, co ktoś inny pomyślał, Keczua używają fraz, które można przetłumaczyć na „Co by na to powiedział?" zamiast „Co by pomyślał?" lub na przykład „powiedzieć nie" zamiast „odmówić". Wydaje się więc, że życie społeczne Keczua jest wyjaśniane przy pomocy całkowicie behawiorystycznych wzorów, co stanowi przesłankę ku temu, że psychologia potoczna w ogóle, nie tylko psychologia potoczna oparta o teorię umysłu, nie jest zjawiskiem uniwersalnym kulturowo.

5. Implicytna teoria umysłu?

W obliczu przedstawionej powyżej różnorodności kulturowej w psychologiach potocznych, zwolennicy teorii teorii są zmuszeni do postulowania, że pomimo tych różnic widocznych na poziomie kulturowym, członkowie każdej kultury posiadają i aplikują implicytną teorię umysłu. Niezależnie od tego, w jaki sposób dana kultura tłumaczy swoje zrozumienie zachowań ludzkich, podświadomie jest ono zawsze podszyte teorią umysłu i implicytnymi inferencjami. Jednak i ta teza jest postawiona pod znakiem zapytania w obliczu wyników badań psychologii rozwojowej przeprowadzonych w szeregu kultur.

Przedstawiona wcześniej panująca na Samoa psychologia potoczna odbija się na wynikach testów fałszywych przekonań przeprowadzonych wśród tam-

tejszych dzieci. Badania Mayera i Trauble'a z 2012 pokazały, że samoańskie dzieci zdają test fałszywych przekonań średnio w wieku około 8 lat (dla dzieci z Zachodu jest to 4 lata), a część z nich dopiero w wieku lat 12 (Mayer, Trauble 2012: 26). Jak argumentują autorzy badań, nawet przy uwzględnieniu innych możliwych czynników, obserwowane opóźnienie jest najprawdopodobniej wynikiem panującej na wyspach psychologii potocznej i ideologii języka.

Również w Japonii różnica w aparacie pojęciowym skutkuje różnicą w wyjaśnieniach podawanych przez dzieci w testach fałszywych przekonań, które z kolei prowadzą do tego, że japońskie dzieci nie zdają tych testów w tym samym okresie co zachodnie dzieci, nie ze względu na to, że nie rozumieją mentalnego świata, ale dlatego, że test mierzy jego zrozumienie jedynie za pomocą pojęć teorii umysłu specyficznej dla Zachodu (w niektórych wersjach testu dzieci są punktowane jedynie za odnoszenie się do umysłu i stanów mentalnych w wyjaśnieniach swoich wyborów).

Badania przy użyciu tak zwanej skali teorii umysłu również pokazują znaczące różnice w kolejności zdawanych przez dzieci zadań wchodzących w jej skład. Skala teorii umysłu jest narzędziem opracowanym przez Wellman, Liu (2004) i ma na celu badanie sekwencji rozwoju poszczególnych zdolności poznania społecznego przy pomocy szeregu zadań. Do badanych zdolności należą zrozumienie pragnień, czy klasyczny test fałszywych przekonań. Wellman *et al.* (2006; 2011) ustalili różnice w kolejności zdawania dwóch elementów tej skali pomiędzy dziećmi z Chin i Iranu a dziećmi z Zachodu. Zachodnim dzieciom przychodzi łatwiej zrozumieć, że różni ludzie mogą mieć różne, ale wciąż prawdziwe przekonania (jest to oddzielny test od testu fałszywych przekonań, ogólnie dzieci dużo lepiej radzą sobie z rozumieniem prawdziwych przekonań) niż to, że brak kontaktu zmysłowego ze stanem rzeczy prowadzi do ignorancji na jego temat. Wśród dzieci z Chin i Iranu kolejność ta jest odwrócona. Dużo bardziej znaczącą różnicę ustalili niedawno Dixson *et al.* (2017), którzy zaaplikowali rozszerzoną skalę teorii umysłu w Vanuatu, kulturze Pacyfiku, gdzie podobnie jak na Samoa panuje doktryna o nieprzeniknioności innych umysłów. Otrzymali oni sekwencję znacząco odmienną od zachodniej, szczególnie wśród grupy badanych pochodzących z obszarów wiejskich; na przykład test badający zrozumienie, że osoba może ukrywać swoje prawdziwe emocje, był dla zachodnich dzieci (chińskich i irańskich również) najtrudniejszym na skali, dla dzieci z Vanuatu był jednym z łatwiejszych. Sam test fałszywych przekonań dzieci z Venuatu zdawały z dużym opóźnieniem relatywnie do Zachodu – podobnie jak te z Samoa – średnio około 12 roku życia.

Bezpośrednio związane z omawianą kwestią są badania nad wpływem języka i interakcji międzyludzkiej na rozwój teorii umysłu. Kultura, wraz z psychologią potoczną w niej panującą, jest przekazywana nowym jej członkom przez innych ludzi poprzez z nimi interakcję, głównie za pomocą języka; stąd natura interakcji społecznej, a szczególnie interakcji językowej staje się ważnym

obiektem badań. Takie aspekty rozwoju jak język używany w domu, „prywatne epistemologie" rodziców, interakcja z rodzeństwem, specyfika środowiska językowego dzieci niesłyszących oraz inne czynniki powiązane z pragmatyką języka są bezpośrednio związane z omawianą tematyką: jeśli czynniki kulturowe wpływają na rozwój teorii umysłu, psychologowie są zainteresowani tym, w jaki konkretnie sposób kultura oddziałuje na jej rozwój, oraz tym, które z aspektów środowiska kulturowego są szczególnie znaczące w tym procesie.

Obecnie istnieje już dosyć pokaźna bateria badań dotyczących powyższych kontekstów i ogólne konkluzje współgrają z obecnymi rozważaniami na temat psychologii potocznej; rodzaj języka otaczający dziecko – sposób wyrażania się o życiu mentalnym, który je otacza – bezpośrednio odbija się na jego pojęciowym rozwoju, w tym rozwoju teorii psychologicznych. Przykład prawdopodobnie najlepiej to obrazujący pochodzi z badań nad wyłaniającym się językiem migowym wśród grupy osób niesłyszących w Nikaragui (Pers, Senghas 2009). Nowy język był początkowo ubogi w wyrażenia i pojęcia odnoszące się do życia mentalnego. Osoby, które posługiwały się tą wczesną formą języka, nie były w stanie zdać standardowego testu fałszywych przekonań nawet w dorosłości. Te same osoby, po przyswojeniu rozbudowanej o pojęcia mentalne wersji języka nabyły zdolność myślenia o fałszywych przekonaniach i w późniejszym punkcie badań zdały test. Więcej przykładów z innych kultur w Strijbos, Bruin (2013) oraz Gut, Mirski (2016).

W obliczu powyższych wyników teza o uniwersalności implicytnej teorii umysłu, która leży u podstaw pozornych różnic w psychologiach potocznych, staje się zagrożona. Różnice w zdroworozsądkowym myśleniu o innych umysłach odbijają się nie tylko w eksplicytnych wyjaśnieniach, ale również w zdolnościach poznawczych jej członków, badanych przez przedstawione testy psychologiczne, które kontrolują poziom rozwoju językowego badanych. Jeśli teoria umysłu jako struktura poznawcza jest, tak jak ujmują ją teoretycy teorii, wrodzonym, uniwersalnym i niezależnym od środowiska kulturowego elementem ludzkiej psychiki, to dlaczego ludzie pochodzący z różnych kultur wykazują tak znaczące różnice w zdawaniu testów psychologicznych badających zrozumienie życia mentalnego? Poniżej przedstawiam dwa główne sposoby odpowiedzenia na to pytanie oferowane przez teoretyków teorii.

6. Dwa wyjścia teoretyczne dla zwolenników teorii teorii

W 2005 Onishi i Baillargeon zademonstrowały, że 15-miesięczne dzieci wykazują zaskoczenie, gdy obserwowana osoba załamuje swoje fałszywe przekonanie. Wykorzystana przez autorów badania metodologia dyshabituacyjna polega na założeniu, że dziecko patrzy się dłużej na sytuację, którą uważa za nietypową. Dzieci badane w tym *niewerbalnym* teście fałszywych przekonań patrzyły się

dłużej, kiedy obserwowana osoba posiadająca fałszywe przekonanie kierowała się do pojemnika, gdzie zabawka znajdowała się naprawdę. Kolejne badania wydają się dawać podobne wyniki nawet wśród 7-miesięcznych niemowląt (Kovács *et al.* 2010; Scott i Baillargeon 2009; Surian i Geraci 2012), choć pojawiają się też wątpliwości co do ich replikowalności (Kulke i Rakoczy 2018). Niezależnie od tych wątpliwości teoretycy teorii skwapliwie przyjmują te wyniki jako argument za tezą o wrodzoności teorii umysłu. Niewerbalne czytanie w umyśle istotnie wzmacnia pozycję teorii teorii, ale nie rozwiązuje samo w sobie problemu międzykulturowej różnorodności w potocznych psychologiach i werbalnych testach na przestrzeni kultur. Podejmując się tego zadania, teoretycy teorii przedstawiają poniższe rozwiązania.

a. Dwa systemy

Podejście dwusystemowe jest najpopularniejszą próbą rozwiązania problemów wynikłych z badań międzykulturowych (Apperly, Butterfill 2009; Apperly, Ian 2012; Low *et al.* 2016; Wellman 2014; por. Strijbos, Bruin 2013). Najogólniej rzecz ujmując, zakłada się tutaj, że „proste", „trzonowe" zrozumienie stanów mentalnych jest dostarczane przez wrodzony mechanizm, który rozwija się już we wczesnym etapie życia (niemowlęctwo) i dostarcza zdolności poznawczych o ograniczonej formie, wystarczającej jednak do zdania implicytnego testu fałszywych przekonań. Ten wrodzony system 1, którego funkcjonowanie obserwujemy w niewerbalnych badaniach nad rozumieniem fałszywych przekonań u niemowląt, jest uniwersalny i niezależny od czynników kulturowych. System 2 natomiast jest systemem nabytym kulturowo i pomimo tego, że w pewnym sensie nabudowuje się na systemie 1, to jest konstytuowany przez pojęcia specyficzne dla danej kultury. To różnice w tym systemie obserwujemy, badając różne kultury przy pomocy testów eksplicytnego zrozumienia stanów mentalnych. Podejście dwusystemowe ma na celu zachowanie tezy o uniwersalizmie podstawowego zrozumienia stanów mentalnych, przy jednoczesnym wyjaśnieniu różnic kulturowych. W podejściach tego typu przyjmuje się, że system 2 w pewnym momencie rozwoju staje się systemem dominującym i to za jego pomocą podmiot wyjaśnia i przewiduje zachowania innych; zrozumienie płynące z wrodzonego systemu 1 jest wciąż implicytnie obecne, ale bierze udział w psychologii potocznej jedynie pomocniczo lub wcale.

Rozbijając czytanie w umyśle na dwa systemy, rozwiązuje się więc tutaj dwa główne problemy interpretacyjne współczesnych badań, jednocześnie zachowując pierwotne założenia wrodzoności teorii umysłu. Po pierwsze, takie postawienie sprawy wyjaśnia różnice kulturowe – są one wynikiem różnic w systemie 2, a nie we „właściwej" teorii umysłu, która skrywa się w systemie 1. Dzieci robią błędy w eksplicytnych eksperymentach, ponieważ testy te wymagają werbalnych odpowiedzi, co toruje (*primes*) użycie systemu 2, lub dlatego, że system 2

zaczyna być dominującym w miarę jego akwizycji i na pewnym etapie rozwoju dzieci zaczynają aplikować go w pierwszej kolejności pomimo tego, że nie jest on jeszcze w pełni rozwinięty. W konsekwencji otrzymuje się tym samym wyjaśnienie, dlaczego już małe dzieci zdają implicytne testy fałszywych przekonań, a eksplicytne, tj. werbalne, zdają dopiero po 4 roku życia: system 1 jest odpowiedzialny za wyniki w testach niewerbalnych i działa bez zarzutu od początku życia, natomiast system 2, który rozwija się wolniej, jest odpowiedzialny za wyniki w testach werbalnych i osiąga wystarczającą sprawność dopiero w wieku lat 4 na Zachodzie, a w innych kulturach później.

b. Jeden system i czynniki pragmatyczne

Teoria modularno-natywistyczna, którą można nazwać teorią jednego systemu, przyjmuje nieco inną linię argumentacyjną w obliczu różnic międzykulturowych. Od pojęcia dwusystemowego różni się tym, że zakłada się tutaj, że wrodzony moduł czytania w umyśle, niezależnie od kulturowo-specyficznej teorii psychologii potocznej, zawsze stoi za zrozumieniem zachowania innych. Wrodzony repertuar konceptualny może być ubogacany przez kulturowo-specyficzne pojęcia, ale nie zachodzi tutaj żadna jakościowa zmiana aplikowanej teorii, tak jak ma to miejsce w założeniach podejścia dwusystemowego (Carruthers 2013, 2016, 2017). Pozostawia to otwartą problematyczną kwestię różnic międzykulturowych, z którymi poprzednia linia argumentacyjna radziła sobie, postulując dwa jakościowo różne systemy.

Teoria modularno-natywistyczna przyjmuje inną strategię i odwołuje się do kwestii pragmatycznych. Argumentuje się tutaj, że dzieci w różnych kulturach mogą mieć lepiej lub gorzej rozwinięte funkcje wykonawcze oraz rozumieć zadania testowe na inne sposoby w zależności od kultury (Westra, Carruthers 2017). Kultura ma więc wpływać na rozwój innych zdolności poznawczych – takich jak kontrola inhibicyjna czy rozumienie implikatur języka – ale na samą teorię umysłu, poza jej ubogacaniem o nowe pojęcia, nie ma wpływu. Dziecko rozpoczyna więc w tym ujęciu życie z systemem 1, a późniejsze problemy w zadaniach werbalnych nie są wynikiem zmiany w sposobie myślenia, tak jak proponują zwolennicy dwóch systemów, a jedynie kwestii wykonawczych.

7. Problemy teorii teorii

W odniesieniu do powyższych interpretacji oraz do teorii teorii w ogóle, można wysnuć przynajmniej trzy poważne zarzuty: a) wyniki testów dyshabituacyjnych na niemowlętach można interpretować na szereg innych sposobów, niż robi się to w teorii teorii. Sam postulat podświadomego wnioskowania budzi kontrowersje, jako że wnioskowania są zazwyczaj postrzegane jako świadomy

proces wysokiego rzędu poznawczego (Newen 2015: 5); b) pojęcie wrodzoności, niemal bezkrytycznie stosowane w teorii teorii, jest źródłem kontrowersji w ogólnym sensie. Z biologicznego punktu widzenia poza zamętem nie wnosi nic do wyjaśnień ontogenetycznych. Ponadto wrodzoność pojęć jest szczególnie problematyczna, jako że wymagałaby ukształtowanych struktur neuronalnych będących korelatami tych pojęć na samym początku życia – jest to sprzeczne z tym, co wiemy o systemie nerwowym; c) paradygmat reprezentacyjno-komputacyjny, w ramach którego ujmuje się ludzkie poznanie w teorii teorii, rodzi poważne teoretyczne problemy w analizach rozwojowych. Poniżej przedstawiam pokrótce te trzy linie argumentacyjne.

a) Teoria teorii bazuje na założeniu, że zdolność czytania w umyśle jest umożliwiona dzięki wrodzonej teorii umysłu (na wrodzonej wiedzy o treści propozycjonalnej). Jednakże tylko werbalne testy dają pewność, że badany podmiot posiada zrozumienie innych umysłów w kategoriach przypisywania im postaw propozycjonalnych, tj. opiera się nie tylko na teorii, ale na teorii umysłu. Taką pewność mamy w odniesieniu do kulturowo zróżnicowanego systemu drugiego; w przypadku uniwersalnego i wrodzonego systemu 1 sprawa jest już dużo bardziej problematyczna. Surowe dane empiryczne są jedynie takie, że niemowlęta patrzą dłużej na sytuację, w której podmiot działa wbrew swoim fałszywym przekonaniom. Zdolność podszywająca te wyniki może być jednak innej natury niż teoretyczna: może być behawiorystyczną teorią (Bartsch *et al.* 2014; Ruffman, Taumoepeau 2014; Perner, Ruffman 2005) lub procesem niższego poziomu poznawczego, który w niepojęciowy sposób przechowuje informacje na temat najbliższego kontekstu sytuacyjnego (Heyes 2014; por. Bruin, Newen 2012). Dziecko nie musi używać pojęć przekonania i pragnienia, żeby zachowywać się w sposób, w jaki obserwujemy w tych eksperymentach.

b) Krytyka pojęć wrodzoności oraz adaptacyjnych wyjaśnień struktur psychologicznych jest przedstawiona w pierwszej kolejności w pracach Lewontina, Goulda czy Oyamy (Gould, Lewontin 1979; Oyama *et al.* 2000; Lewontin 2001). W etologii, na przykład, postrzega się już pojęcie wrodzoności jako nieprzydatne – zachowanie jest wynikiem procesu rozwojowego, w którym grają rolę nie tylko czynniki genetyczne czy wewnątrzorganizmowe, ale również pozaorganizmowe, czyli środowiskowe. Wpływ poszczególnych elementów takiego systemu rozwojowego będzie oczywiście różny dla poszczególnych zachowań na różnych etapach rozwoju osobniczego, ale czołową obserwacją jest tutaj to, że żadna cecha fenotypiczna nie preegzystuje na poziomie genetycznym – geny są tylko jednym elementem złożonego, „wielokontygencyjnego" (*multiply contigent*) systemu przyczynowego (choć oczywiście niezbędnym). Często bywa tak, że czynnik genetyczny wymaga obecności czynnika środowiskowego, aby mieć znaczenie rozwojowe, a następnie wykształcona w ten sposób cecha fenotypiczna umożliwia interakcję kolejnego czynnika genetycznego z kolejnym elementem środowiska itd.

Mowa o wrodzoności jest więc źródłem nieporozumień, jako że nie jest jasne, co tak naprawdę to pojęcie ma znaczyć, szczególnie w analizach ontogenetycznych. Fakt, że dana cecha była obiektem naturalnej selekcji w filogenezie, nie oznacza, że jest ona automatycznie obecna w ontogenezie – teoria rozwojowa musi być w stanie podać, w jaki sposób ewolucyjna spuścizna wchodzi w interakcję ze środowiskiem rozwoju i prowadzi do wykształcenia się danej cechy w organizmie. Wyjaśnienie rozwojowe *per se* ma na celu przedstawić wpływ wszystkich niezbędnych czynników przyczynowych prowadzących do wyłonienia się danej cechy, opisać zmiany systemu rozwojowego na przestrzeni czasu; stwierdzanie „wrodzoności" cechy nie jest w żadnym wypadku wyjaśnieniem, a jedynie uniknięciem odpowiedzi. Innymi słowy, teorie rozwojowe, które powołują się na wrodzoność, biorą *explanandum* za *explanans*, są wyjaśnieniem typu *virtus dormitiva*.

Postulat wrodzonej teorii umysłu podpada pod ten błąd metodologiczny; oferuje się tu jako wyjaśnienie zdolności poznania społecznego wrodzony zestaw pojęć, którego ontogenezę zrzuca się na barki biologii. Jednak w biologii rozwojowej, jak zauważyliśmy, nie ma miejsca na pojęcie wrodzoności – rozwój uwzględnia czynniki niebiologiczne tak samo jak biologiczne, środowiskowe tak samo jak nieśrodowiskowe. Mimo że nie ma nic złego w budowaniu teorii na wyższym poziomie analizy, gdzie przyjmuje się pewne zdania za aksjomaty wyjaśnialne z poziomu nauki niższej, to przyjmując takie zdanie musimy być pewni, że w rzeczy samej funkcjonuje ono na niższym poziomie. Jeśli tak nie jest, skazujemy się na proponowanie wyjaśnień w teoretycznej próżni, niezgodnej z nauką, na której bazujemy. Jak piszą Lewis *et al.* (2013), „nie można zaprojektować *perpetuum mobile*, a potem zrzucać problem jego konstrukcji na barki inżyniera. Potrzebujemy jasno wykazać, co mamy na myśli, używając pojęcia «wrodzone», a nie odsyłać do biologa, mówiąc, że to już nie mój wydział".

Nieużyteczność pojęcia wrodzoności nie odnosi się oczywiście do postulatu, że poznanie społeczne jest oparte na teorii umysłu, a jedynie do postulatu jej wrodzoności. Jest wciąż pytaniem otwartym, czy już 7-miesięczne dzieci posługują się teorią umysłu, choć samo stwierdzenie, że jest to teoria wrodzona, nie wyjaśnia niczego. Jest również możliwe, że wykształcenie się tej teorii zachodzi poprzez interakcje czynników genetycznych i innych czynników wewnątrzorganizmowych – bez wpływu czynników społecznych, kulturowych, czy nawet ekologicznych, czyli tak jak teoretycy teorii prawdopodobnie mają na myśli, używając pojęcia wrodzoności. Wydaje się to jednak wątpliwe, biorąc pod uwagę, jak wrażliwy jest rozwój zdolności poznania społecznego na czynniki społeczne i kulturowe w późniejszych etapach życia.

Ponadto istnienie takich „wrodzonych" pojęć, czyli uformowanych niezależnie od doświadczenia w środowisku, jest sprzeczne z naszym zrozumieniem układu nerwowego. Jeśli przyjmiemy najbardziej podstawową definicję fizykalną pojęcia jako wzoru aktywacji neuronalnej o określonej funkcji przyczynowej,

to wrodzone pojęcie będzie musiało być takim z góry ustanowionym schematem aktywacji. System nerwowy jest jednak ogromnie plastyczny i otwarty na diametralne zmiany na początku życia człowieka i stąd „zaprogramowane" schematy aktywacji przy narodzinach są tezą trudną do obrony, szczególnie w korze nowej mózgu, gdzie znajdujemy korelaty neuronalne pojęć (Elman 1996). Faktem jest, że naturalna selekcja odgrywa znaczącą role w uformowaniu fenotypu poprzez wpływ na geny, jednak mikromorfologia, włączając połączenia neuronalne, jest raczej nastrojona na wpływ środowiska (Pezzulo *et al.* 2015), co ma również sens z adaptacyjnego punktu widzenia, gdyż umożliwia organizmowi większą sprawność w nowym środowisku, nawet jeśli jest ono odmienne od środowiska przeszłych pokoleń. Teoria umysłu jako struktura psychologiczna rozwojowo niezależna od środowiska jest więc mało prawdopodobną tezą.

c) Reprezentacyjno-komputacyjny paradygmat w kontekście rozwojowym spotyka się z również krytyką formalną. Allen i Bickhard (2013, 2011) zarzucają fundacjonalizm teoriom reprezentacyjno-komputacyjnym – których teoria teorii jest przykładem – wykazując, że pogląd ten jest nie tylko problematyczny ze względu na pojęcie wrodzoności, które w ujmowaniu rozwoju poznawczego się tam przyjmuje, ale również ze względu na ograniczenia czysto teoretyczne, które *a priori* uniemożliwiają postawienie pewnych, w innym przypadku prawdopodobnych, hipotez.

Po pierwsze, paradygmat reprezentacyjno-komputacyjny opiera się na założeniu, że działanie organizmu zawsze musi opierać się na podświadomych reprezentacyjnych procesach poznawczych, stanowisko sięgające daleko wstecz kognitywistyki, ale chyba najwydatniej wyartykułowane przez Fodora (Fodor 1975, 2010; Cain 2013). W połączeniu z założeniem, że percepcja jest raczej uboga w dane niepowiązane bezpośrednio z modalnościami zmysłowymi (zob. Gallagher 2008), prowadzi to do potrzeby postulowania (implicytnych) teorii wszędzie tam, gdzie obserwujemy zachowanie w swojej inteligencji wykraczające poza to, co jest percypowane. W rezultacie otrzymujemy cały zestaw teorii zakładających „trzonową wiedzę" dotyczącą dziedzin, w których niemowlęta są nad wyraz biegłe: fizyki, gramatyki, matematyki, biologii oraz oczywiście psychologii (Spelke, Kinzler 2007; Spelke 2003, 1994; Pinker 2014/1994; Pinker 1994; Wynn 1992). Przyjmuje się pojęciowe reprezentacje umysłowe wchodzące w skład tych wrodzonych teorii jako bazę dla dalszego rozwoju poznawczego, wykluczając w efekcie możliwość, że działanie rozwija się najpierw i opiera się na procesach niereprezentacyjnych, albo przynajmniej niepojęciowych, następnie prowadząc do wykształcenia się mentalnych reprezentacji o pojęciowym charakterze.

Prawdziwość tej hipotezy jest kwestią otwartą, ale problem leży w tym, że dla reprezentacyjno-komputacyjnej teorii jest niemożliwym nawet jej postawienie ze względu na ograniczenia teoretyczne. Potrzeba reprezentacjonalistycznego wyjaśniania zachowań jest powodem, dla którego teoretyk teorii wydaje się

nie mieć innego wyjścia jak przypisywać teorię umysłu 7-miesięcznym dzieciom zdającym niewerbalny test fałszywych przekonań, z góry wykluczając alternatywne interpretacje o bardziej mechanistycznym charakterze, o których była mowa w pkcie (a). Mechanistyczne wyjaśnienia procesów poznawczych leżą poza możliwościami teoretycznymi paradygmatu reprezentacyjno-komputacyjnego, a jest możliwym, że takiego właśnie wyjaśnienia potrzebujemy.

Co więcej, fundacjonalistyczne ujęcie rozwoju poznawczego jest problematyczne ze względu na to, że wyklucza możliwość autentycznego uczenia się nowych pojęć. Jako że reprezentacyjne zjawiska nie mogą wykształcić się z niereprezentacyjnych, treść nowych pojęć jest jedynie pozyskiwana z wcześniejszych i ten nieskończony regres kończy się na wrodzonych pojęciach. Idealnym przykładem tego problemu jest wyjątkowo płodna metodologia badań modelująca dziecięce teorie przy pomocy statystyki Bayesowskiej, gdzie dziecko ma nabywać nowe pojęcia poprzez formułowanie hipotez i testowanie ich w doświadczeniu (Gopnik, Bonawitz 2015; Gopnik, Wellman 2012; Perfors *et al.* 2011; Gopnik, Tenenbaum 2007): pomimo ogromnych sukcesów empirycznych nadal zakłada się tutaj możliwość sformułowania pierwszej hipotezy, do czego potrzeba początkowych pojęć, co wymusza fundacjonalizm i natywizm (zob. Bickhard 2016).

Kolejnym problematycznym ruchem argumentacyjnym przyjmowanym często wśród zwolenników teorii dziecięcych teorii jest rozgraniczenie pomiędzy kompetencją a wykonaniem (*competence – performance*). Dotyczy ono przede wszystkim jednosystemowej wersji teorii teorii (zob. 6b). Na przykład Westra i Carruthers (2017) argumentują, że czynniki wykonawcze – niezrozumienie implikatur pytań testowych, nierozwinięte funkcje wykonawcze czy inne tendencje poznawcze małych dzieci – stoją za postępami na skali umysłu (zob. 5) a nie nierozwinięte jeszcze zrozumienie przekonań czy pragnień. Carruthers jest znany ze swojego natywistycznego stanowiska (2013, 2015, 2016) i często jego argumentacja przybiera właśnie taki kształt, analogiczny do tego, co chyba najwydatniej podkreślał Chomsky w kontekście języka.

Allen i Bickhard (2013) argumentują przeciwko rozgraniczeniu pomiędzy kompetencją a wykonaniem ujętym w ten sposób: pokazują, że jako rozgraniczenie teoretyczne jest ono eksplanacyjnie bezużyteczne (jako rozgraniczenie metodologiczne jest oczywiście konieczne i pożądane – w zależności od eksperymentu inne zmienne są przyjmowane jako kompetencja a inne jako czynniki wykonawcze). Jak pokazują, rozróżnienie *competence – performance* błędnie spaja opis z wyjaśnieniem: opis zjawiska, które obserwujemy i stajemy przed zadaniem jego wyjaśnienia, jest przemianowywany w postaci bardziej abstrakcyjnego mechanizmu, który domniemanie wyjaśnia to zjawisko. I tak, w przypadku języka generatywne zdolności gramatyczne użytkowników są „wyjaśnione" przy pomocy wrodzonego mechanizmu czy wiedzy gramatycznej, a w przypadku poznania społecznego zdolność do rozumienia stanów mentalnych in-

nych ludzi jest „wyjaśniona" w kategoriach wrodzonego mechanizmu, który umożliwia rozumienie tych stanów. Po raz kolejny mamy do czynienia z błędem logicznym typu *virtus dormitiva*, który miesza to, co ma być wyjaśnione, z wyjaśnieniem. Zaobserwowane zjawisko rozumienia innych przy pomocy teorii umysłu jest awansowane do statusu mechanizmu poznawczego, który następnie postrzegany jest jako funkcjonujący w umyśle od samego początku życia, a problemy z jego aplikacją widzi się jako problemy jedynie z wykonaniem – to inne czynniki, czynniki performatywne, stoją na drodze ekspresji zrozumienia płynącego z tego mechanizmu. Pomimo swojego pozornego sensu wyjaśnienie takie nie wyjaśnia w gruncie rzeczy nic, a jedynie powtarza to, co ma być wyjaśnione w bardziej abstrakcyjny sposób: jeśli zdolność do czytania w umyśle ma być wytłumaczona przez wrodzoną teorię umysłu, która umożliwia postrzeganie rzeczywistości w kategoriach mentalistycznych, to wciąż stoimy przed zadaniem wyjaśnienia, skąd i jak ta zdolność się rozwija.

8. Alternatywne teorie

Na podstawie powyższego przeglądu problemów, na jakie napotyka dominująca w literaturze teoria teorii, jesteśmy w stanie wyłonić czołowe wyzwania stojące przed potencjalnymi alternatywami. Po pierwsze, teoria zajmująca się rozwojem psychologicznym musi być zgodna ze współczesną wiedzą na temat zjawiska rozwoju, ostrożnie używać pojęcia wrodzoności i wykazać, w jaki sposób reprezentacje mentalne wyłaniają się ze zjawisk niereprezentacyjnych *w ontogenezie* lub dostarczyć dużo bardziej przekonujących dowodów, że następuje to w filogenezie. Szereg psychologów i filozofów sugeruje inne ramy teoretyczne – teorię systemów rozwojowych (*developmental systems theory*) – aby adekwatnie opisać i wyjaśnić zjawiska rozwojowe (Nelson 2007, 2005; Oyama *et al.* 2000; Lewis *et al.* 2013). Jeśli chodzi o ontogenetyczny rozwój reprezentacji umysłowych, teorie oparte na działaniu (*action-based theories*) skupiają coraz więcej uwagi (Engel *et al.* 2015); są one zgodne z teorią systemów rozwojowych oraz umożliwiają wyjaśnianie rozwoju reprezentacji umysłowych poprzez postulowanie, że działanie poprzedza reprezentacje umysłowe. Przykładami takich podejść są interaktywizm Bickharda (Bickhard 2009b, 2009a; Allen i Bickhard 2011b, 2011a), społeczność umysłów (*community of minds*) Nelson (Nelson 2007, 2005), społeczno-komunikatywne podejście (*social-communicative approach*) Carpendale'a i Lewisa (Carpendale, Lewis 2006, 2004, 2012) czy podejście Gallaghera skupiające się na bezpośrednio postrzeganej intersubiektywności (Gallagher 2012; Gallagher, Povinelli 2012; Gallagher 2008; Gallagher, Hutto 2008).

Z racji tego, że przypisuje się w tych podejściach dużo większą konstytutywną siłę działaniu i doświadczeniu konsekwencji swoich działań w środowisku,

całkiem inaczej podchodzi się tam do przedstawionych różnic międzykulturowych w czytaniu w umyśle, niż ma to miejsce w teorii teorii. Niereprezentacyjne wyjaśnienie zachowania niemowląt w niewerbalnym teście fałszywych przekonań również jest z racji tego umożliwione. Wyczerpująca analiza tych podejść leży jednak poza zakresem obecnego artykułu, chciałem je jedynie zasygnalizować jako kierunek przyszłych rozważań.

9. Wnioski

Moim celem było przedstawienie problemów teorii teorii w obliczu różnic kulturowych w potocznych psychologiach, z uwzględnieniem wyników badań poznania społecznego przeprowadzonych z dziećmi pochodzącymi z tych kultur. Teoria teorii rozgranicza teorię umysłu rozumianą jako rozwojowo pierwotną strukturę poznawczą, umożliwiającą poznanie społeczne oraz teorię umysłu, rozumianą jako eksplicytna teoria psychologii potocznej. Na pierwszy rzut oka umożliwia to postulat, że różnorodność międzykulturowa w psychologiach potocznych jest jedynie eksplicytną formą implicytnie uniwersalnej teorii umysłu jako struktury poznawczej. Wyniki badań psychologicznych nad poznaniem społecznym w kulturach o różnych psychologiach potocznych utrudniają jednak taką interpretację: potoczna psychologia wpływa na zdolność myślenia o innych umysłach mierzoną w eksperymentach niezależnie od języka. W obliczu tych wyników teoretycy teorii uciekają się do jednego z dwóch wyjść teoretycznych: teorii dwóch systemów lub teorii jednego systemu. Jednak obydwa te wyjścia postulują wrodzoną wiedzę, różniąc się jedynie ilościowo względem tego, ile z tej wiedzy jest wrodzone, i w konsekwencji stoją przed dalszymi wyzwaniami teoretycznymi: problematycznym podejściem do wrodzoności, fundacjonalizmem, który apriorycznie uniemożliwia stawianie pewnych hipotez, oraz innymi problemami pojęciowymi, takimi jak rozgraniczenie pomiędzy kompetencją a wykonaniem. Wielu naukowców sugeruje potrzebę zmiany paradygmatu teoretycznego, aby rozwiązać te problemy, sugerując alternatywy, które zasygnalizowałem na końcu artykułu.

Bibliografia

Allen, J.W.P., Bickhard, M.H. (2011a). Emergent Constructivism. *Child Development Perspectives* 5 (3), 164-165. DOI: 10.1111/j.1750-8606.2011.00178.x.

Allen, J.W.P., Bickhard, M.H. (2011b). Normativity. A Crucial Kind of Emergence. *Human Development*, 54 (2), 106-112. DOI: 10.1159/000327096.

Allen, J.W.P., Bickhard, M.H. (2013). Stepping off the pendulum. Why only an action-based approach can transcend the nativist-empiricist debate. *Cognitive Development*, 28 (2), 96-133. DOI: 10.1016/j.cogdev.2013.01.002.
Apperly, I.A. (2012). Mindreaders. The cognitive basis of 'theory of mind'. Hove: Psychology.
Apperly, I.A., Butterfill, S.A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological review* 116 (4), 953-970. DOI: 10.1037/a0016923.
Bickhard, M.H. (2009a). Interactivism. W: J. Symons, P. Calvo (red.), *The Routledge companion to philosophy of psychology Calvo* (s. 345-359). London–New York: Routledge Taylor & Francis Group.
Bickhard, M.H. (2009b). The interactivist model. *Synthese*, 166 (3), 547-591. DOI: 10.1007/s11229-008-9375-x.
Bickhard, M.H. (2016). Probabilities over What? *Human Development*, 59 (1), 34-36. DOI: 10.1159/000447026.
Bruin, L.C. de; Newen, A. (2012). An association account of false belief understanding. *Cognition*, 123 (2), 240-259. DOI: 10.1016/j.cognition.2011.12.016.
Cain, M.J. (2013). Fodor. Language, Mind and Philosophy. Hoboken: Wiley (Key Contemporary Thinkers), dostępne na stronie internetowej (06.04.2018): http://gbv.eblib.com/patron/FullRecord.aspx?p=1215760.
Carpendale, J.I.M., Lewis, Ch. (2004). Constructing an understanding of mind. The development of children's social understanding within social interaction. *The Behavioral and brain sciences*, 27 (1), 79-96 (discussion 96-151).
Carpendale, J.I.M., Lewis, Ch. (2006). How children develop social understanding. Malden, Mass., Oxford: Blackwell Publishers.
Carpendale, J.I.M., Lewis, Ch. (2012). Reaching, requesting and reflecting. From interpersonal engagement to thinking. W: A. Foolen *et al.* (red.), *Moving ourselves, moving others. Motion and emotion in intersubjectivity, consciousness and language*. Amsterdam–Philadelphia: John Benjamins Pub. Co.
Carruthers, P. (2013). Mindreading in Infancy. *Mind & Language*, 28 (2), 141-172. DOI: 10.1111/mila.12014.
Carruthers, P. (2016). Two Systems for Mindreading? *Review of Philosophy and Psychology*, 7 (1), 141-162. DOI: 10.1007/s13164-015-0259-y.
Carruthers, P. (2017). Mindreading in adults. Evaluating two-systems views. *Synthese*, 194, 3, 673-688. DOI: 10.1007/s11229-015-0792-3.
Churchland, P.M. (1998). Folk psychology. W: P.M. Churchland, P.S. Churchland (red.), *On the contrary. Critical essays. 1987-1997* (s. 3-15). Cambridge, Mass.: MIT Press (Bradford Bks).
Duranti, A. (2008). Further Reflections on Reading Other Minds. *Anthropological Quarterly*, 81 (2), 483-494. DOI: 10.1353/anq.0.0002.
Engel, A.K., Friston, K.J., Kragic, D. (red.) (2015). *The pragmatic turn. Toward action-oriented views in cognitive science*. Cambridge, Mass.: The MIT Press.
Fodor, J.A. (1975). *The language of thought*. New York: Crowell.

Fodor, J.A. (1983). *The Modularity of Mind*. Cambridge: The MIT Press; dostępne na stronie internetowej: http://gbv.eblib.com/patron/FullRecord.aspx?p=3339779 (06.04.2018).

Fodor, J.A. (1992). A theory of the child's theory of mind. *Cognition*, 44 (3), 283-296. DOI: 10.1016/0010-0277(92)90004-2.

Gallagher, Sh. (2008). Direct perception in the intersubjective context. *Consciousness and cognition*, 17 (2), 535-543. DOI: 10.1016/j.concog.2008.03.003.

Gallagher, Sh., Hutto, D.D. (2008). Understanding others through primary interaction and narrative practice. W: J. Zlatev, T.P. Racine, Ch. Sinha i E. Itkonen (red.), *The Shared Mind* (t. 12, s. 17-38). Amsterdam: John Benjamins Publishing Company.

Gallagher, Sh., Povinelli, D.J. (2012). Enactive and Behavioral Abstraction Accounts of Social Understanding in Chimpanzees, Infants, and Adults. *Review of Philosophy and Psychology*, 3 (1), 145-169. DOI: 10.1007/s13164-012-0093-4.

Gopnik, A., Bonawitz, E. (2015). Bayesian models of child development. *Wiley interdisciplinary reviews. Cognitive science*, 6 (2), 75-86. DOI: 10.1002/wcs.1330.

Gopnik, A., Tenenbaum, J.B. (2007). Bayesian networks, Bayesian learning and cognitive development. *Developmental Science*, 10 (3), 281-287. DOI: 10.1111/j.1467-7687.2007.00584.x.

Gopnik, A., Wellman, H.M. (1992). Why the Child's Theory of Mind Really Is a Theory. *Mind & Language*, 7 (1-2), 145-171. DOI: 10.1111/j.1468-0017.1992.tb00202.x.

Gopnik, A., Wellman, H.M. (2012). Reconstructing constructivism: causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin*, 138 (6), 1085-1108. DOI: 10.1037/a0028044.

Gould, S.J., Lewontin, R.C. (1979). The Spandrels of San Marco and the Panglossian Paradigm. A Critique of the Adaptationist Programme. *Proceedings of the Royal Society B: Biological Sciences*, 205 (1161), 581-598. DOI: 10.1098/rspb.1979.0086.

Gut, A. (2016). Badania kognitywne i rozwojowe nad czytaniem umysłu (mindreading), W: J. Bremer (red.), *Przewodnik po kognitywistyce* (s. 715-758). Kraków: Wydawnictwo WAM.

Heyes, C. (2014). False belief in infancy. A fresh look. *Developmental Science*, 17 (5), 647-659. DOI: 10.1111/desc.12148.

Bartsch K., Young, T.N., Estes, D. (2014). What Young Children Know About Getting a Better Idea. Devil in the Details, W: red. O.N. Saracho, *Contemporary perspectives on research in theory of mind in early childhood education* (s. 69-90), Charlotte, North Carolina: Information Age Publishing.

Kovács, Á.M., Téglás, E., Endress, A.D. (2010). The social sense. Susceptibility to others' beliefs in human infants and adults. *Science*, 330 (6012), 1830-1834. DOI: 10.1126/science.1190792.

Kulke, L., Rakoczy, H. (2018). Implicit Theory of Mind – An overview of current replications and non-replications. *Data in Brief*, 16, 101-104. DOI: 10.1016/j.dib.2017.11.016.

Lewis, Ch., Carpendale, J.I.M., Stack, J. (2013). Anticipation and social interaction. Commentary on "Stepping off the pendulum: Why only an action-based approach can transcend the nativist-empiricist debate" by J. Allen and M. Bickhard. *Cognitive Development*, 28 (2), 159-163. DOI: 10.1016/j.cogdev.2013.01.008.

Lewontin, R.C. (2001). *The triple helix. Gene, organism and environment*. Cambridge Mass.–London: Harvard University Press.

Low, J., Apperly, I.A., Butterfill, S.A., Rakoczy, H. (2016). Cognitive Architecture of Belief Reasoning in Children and Adults. A Primer on the Two-Systems Account. *Child development perspectives*, 10 (3), 184-189. DOI: 10.1111/cdep.12183.

Mameli, M., Bateson, P. (2011). An evaluation of the concept of innateness. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 366 (1563), 436-443. DOI: 10.1098/rstb.2010.0174.

Mayer, A., Träuble, B. (2012). Synchrony in the onset of mental state understanding across cultures? A study among children in Samoa. *International Journal of Behavioral Development*, 37 (1), 21-28. DOI: 10.1177/0165025412454030.

Naito, M. (2014). From Theory of Mind to Theory fo Relation. Sociocultural Perspectives on Japanese Children's Social Understanding. W: O.N. Saracho (red.), *Contemporary perspectives on research in theory of mind in early childhood education* (s. 381-408). Charlotte, North Carolina.

Nelson, K. (2005). Language Pathways into the Community of Minds. W: J.W. Astington, J.A. Baird (red.), *Why language matters for theory of mind New York* (s. 26-49). Oxford: Oxford University Press.

Nelson, K. (2007). *Young minds in social worlds. Experience, meaning, and memory*, Cambridge, Mass., London: Harvard University Press.

Onishi, K.H., Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, 308 (5719), 255-258. DOI: 10.1126/science.1107621.

Oyama, S. (1985/2000). The Ontogeny of Information. Durham: Duke University Press.

Perfors, A., Tenenbaum, J.B., Griffiths, Th.L., Xu, F. (2011). *A Tutorial Introduction to Bayesian Models of Cognitive Development*. Fort Belvoir, VA: Defense Technical Information Center.

Perner, J., Ruffman, T. (2005). Psychology. Infants' insight into the mind: how deep? *Science*, 308 (5719), 214-216. DOI: 10.1126/science.1111656.

Pinker, S. (2014/1994). The language instinct. How the mind creates language. Brilliance Audio.

Premack, D. Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Science*, 1 (04), 515. DOI: 10.1017/S0140525X00076512.

Pyers, J.E., Senghas, A. (2009). Language promotes false-belief understanding: evidence from learners of a new sign language. *Psychological Science*, 20 (7), 805-812. DOI: 10.1111/j.1467-9280.2009.02377.x.

Robbins, J., Rumsey, A. (2008). Introduction. Cultural and Linguistic Anthropology and the Opacity of Other Minds. *Anthropological Quarterly*, 81 (2), 407-420. DOI: 10.1353/anq.0.0005.

Ruffman, T., Taumoepeau, M. (2014). When and How Does a Theory of Mind Arise? W: O.N. Saracho (red.), *Contemporary perspectives on research in theory of mind in early childhood education* (s. 45-63), Charlotte, North Carolina: Information Age Publishing.
Scott, R.M., Baillargeon, R. (2009). Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development*, 80 (4), 1172-1196. DOI: 10.1111/j.1467-8624.2009.01324.x.
Spelke, E.S. (1994). Initial knowledge: six suggestions. *Cognition*, 50 (1-3), 431-445.
Spelke, E.S. (2003). What makes us Smart. Core knowledge and natural language. W: D. Gentner, S. Goldin-Meadow (red.), *Language in mind. Advances in the study of language and thought* (s. 277-311). Cambridge, Mass., London: MIT Press (A Bradford book).
Spelke, E.S., Kinzler, K.D. (2007). Core knowledge. *Developmental Science*, 10 (1), 89-96. DOI: 10.1111/j.1467-7687.2007.00569.x.
Strijbos, D.W., Bruin, L.C. de (2013). Universal belief-desire psychology? A dilemma for theory theory and simulation theory. *Philosophical Psychology*, 26 (5), 744-764. DOI: 10.1080/09515089.2012.711034.
Surian, L., Geraci, A. (2012). Where will the triangle look for it? Attributing false beliefs to a geometric shape at 17 months. *The British journal of developmental psychology*, 30 (1), 30-44. DOI: 10.1111/j.2044-835X.2011.02046.x.
Vinden, P.G. (1996). Junin Quechua Children's Understanding of Mind. *Child Development*, 67 (4), 1707-1716. DOI: 10.1111/j.1467-8624.1996.tb01822.x.
Wellman, H.M. (2014). *Making minds. How theory of mind develops*. Oxford, New York: Oxford University Press.
Wellman, H.M., Fang, F., Liu, D., Zhu, L., Liu, G. (2006). Scaling of Theory-of-Mind Understandings in Chinese Children. *Psychological Science*, 17 (12), 1075-1081. DOI: 10.1111/j.1467-9280.2006.01830.x.
Wellman, H.M., Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, 75 (2), 523-541. DOI: 10.1111/j.1467-8624.2004.00691.x.
Westra, E., Carruthers, P. (2017). Pragmatic development explains the Theory-of-Mind Scale. *Cognition*, 158, 165-176. DOI: 10.1016/j.cognition.2016.10.021.
Wimmer, H., Perner, J. (1983). Beliefs about beliefs. Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13 (1), 103-128. DOI: 10.1016/0010-0277(83)90004-5.
Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, 358 (6389), 749-750. DOI: 10.1038/358749a0.

Streszczenie

W poniższym tekście przedstawiam szereg problemów podejścia do poznania społecznego zwanego teorią teorii w obliczu różnic kulturowych w potocznych psy-

chologiach, z uwzględnieniem wyników badań poznania społecznego przeprowadzonych z dziećmi pochodzącymi z tych kultur. Teoria teorii rozgranicza teorię umysłu rozumianą jako rozwojowo pierwotną strukturę poznawczą, umożliwiającą poznanie społeczne oraz teorię umysłu, rozumianą jako eksplicytna teoria psychologii potocznej. Na pierwszy rzut oka umożliwia to postulat, że różnorodność międzykulturowa w psychologiach potocznych jest jedynie eksplicytną formą implicytnie uniwersalnej teorii umysłu jako struktury poznawczej. Wyniki badań psychologicznych nad poznaniem społecznym w kulturach o różnych psychologiach potocznych utrudniają jednak taką interpretację: potoczna psychologia wpływa na zdolność myślenia o innych umysłach mierzoną w eksperymentach niezależnie od języka. W obliczu tych wyników teoretycy teorii uciekają się do jednego z dwóch wyjść teoretycznych: teorii dwóch systemów lub teorii jednego systemu. Obydwa te wyjścia postulują wrodzony zestaw reprezentacji mentalnych, różniąc się jedynie ilościowo względem tego, ile z tego wrodzonego uposażenia jest wrodzone, a w konsekwencji stoją przed dalszymi wyzwaniami teoretycznymi – problematycznym podejściem do wrodzoności, fundacjonalizmem oraz innymi problemami pojęciowymi, takimi jak rozgraniczenie pomiędzy kompetencją a wykonaniem. Wielu naukowców sugeruje potrzebę zmiany paradygmatu teoretycznego, aby rozwiązać te problemy, sugerując alternatywy, które sygnalizuję na końcu artykułu.

Słowa kluczowe: teoria umysłu, psychologia potoczna, różnice kulturowe, rozwój poznawczy, teoria teorii, poznanie społeczne.

Abstract

In this article, I present a number of problems that the approach to social cognition called „theory theory" encounters in the face of cross-cultural data. Ethnographic data from various cultures show that folk psychology does not always take the form of a theory of mind, and children from such cultures have problems with socio-cognitive tasks of various sorts. Theory theory distinguishes between theory of mind understood as the explicit way of explaining behavior we find in some folk psychologies, and as a sub-personal cognitive structure that enables social cognition and socio-cognitive development. At first glance, this allows the claim that cross-cultural variance in folk psychology (explicit psychological theory) is superficial, with different folk psychologies being mere cultural expressions of an underlying universal theory of mind understood as a cognitive mechanism. Results of psychological experiments conducted within cultures with various folk psychologies speak against such a hypothesis: folk psychology influences the ability to think about other minds as measured in the experiments, which controlled for language. This leads to two further theoretical moves: theory theory proponents either claim that there are two systems – one implicit and one explicit – the first of which is universal while the other is culture-dependent, or they blame performational factors for the cross-cultural variance in performance on the tests. Both of these, however, still commit to

an idea of a representational foundation out of which further socio-cognitive development proceeds; they differ only on how much is given at the outset. This leads to further problems, which I discuss in the text: untenable concept of innateness, developmental foundationalism, as well as other theoretical problems such as the competence-performance distinction. Alternative approaches to social cognition are in play, which I point to by the end of the article.

Keywords: theory of mind, folk psychology, cultural differences, cognitive development, theory theory, social cognition.

Synthese (2020) 197:5511–5537
https://doi.org/10.1007/s11229-018-01976-y

Action-based versus cognitivist perspectives on socio-cognitive development: culture, language and social experience within the two paradigms

Robert Mirski[1] · **Arkadiusz Gut**[1]

Received: 3 September 2017 / Accepted: 5 October 2018 / Published online: 24 October 2018

Abstract
Contemporary research on mindreading or theory of mind has resulted in three major findings: (1) There is a difference in the age of passing of the elicited-response false belief task and its spontaneous–response version; 15-month-olds pass the latter while the former is passed only by 4-year-olds (in the West). (2) Linguistic and social factors influence the development of the ability to mindread in many ways. (3) There are cultures with folk psychologies significantly different from the Western one, and children from such cultures tend to show different timetables of mindreading development. The traditional accounts of the data are nativism, rational constructivism, and two-systems theory. In this paper, we offer criticism of these traditional cognitivist accounts and explore an alternative, action-based framework. We argue that even though they all seem to explain the above empirical data, there are other, theoretical reasons why their explanations are untenable. Specifically, we discuss the problem of foundationalism and the related problem of innateness. Finally, we explore an alternative, action-based framework that avoids these theoretical limitations and offer an interpretation of the empirical data from that perspective.

Keywords Mindreading · Theory of mind · Social cognition · Culture · Development · Action-based · Developmental systems · False-belief task · Folk psychology · Foundationalism · Nativism · Innateness · Interactivism · Constructivism

1 Introduction

In this paper, we offer criticism of traditional cognitivist theories of socio-cognitive development and explore an alternative, action-based framework. The accounts con-

✉ Robert Mirski
robmirski@gmail.com

[1] The John Paul II Catholic University of Lublin, Al. Racławickie 14, 20-950 Lublin, Poland

sidered are nativism, rational constructivism, and two-systems theory. We argue that even though they all seem to explain empirical data about socio-cognitive development—infant mindreading, modulation by experiential factors, and cross-cultural variance—there are other, theoretical reasons why their explanations are untenable. Specifically, we discuss the problem of foundationalism and the related problem of nativism. Finally, we explore an alternative, action-based framework that avoids these theoretical limitations and offer an interpretation of the empirical data from that perspective.

2 Current empirical data

It has been over three decades since Premack and Woodruff's landmark paper that set the course for contemporary research on the human ability to read other minds (Premack and Woodruff 1978; Wimmer and Perner 1983). The area of study, known as theory of mind (ToM) or mindreading, produced a staggering number of empirical findings. Notably, recent years have abounded in significant findings that can be broken down into three groups, which pose a challenge for any theory aiming to account for them:

1. The false belief test has recently been adapted to minimize extraneous cognitive demands on the child.[1] The results from studies adopting this new, *spontaneous–response,* non-verbal FBT are one of the main points of contestation in the field. Infants as young as 15 months pass the spontaneous–response test, as opposed to around four years for the elicited-response version of it (Onishi and Baillargeon 2005; Surian et al. 2007). The methodology is similar, but the crucial difference is that instead of *asking* the child about what she thinks, the child's looking time is measured in both possible scenarios. If the child looks longer at the situation where the observed person violates her false belief, it is interpreted as the child considering this unusual and therefore understanding false beliefs (Träuble et al. 2010). Alternatively, the anticipatory looking of the child is measured before the observed agent makes her choice. If the child passes the spontaneous–response test, she is often considered to possess an "implicit" theory of mind. This finding has been the main line of argumentation for researchers located on the nativist side of the innate-constructed scale.
2. Linguistic and social factors influence the development of explicit theory of mind in many ways (e.g. Astington and Baird 2005; de Villiers and de Villiers 2014; Kristen and Sodian 2014; Milligan et al. 2007; Ruffman et al. 2003). The first inquiries into the significance of language for ToM were trying to rule whether it was any particular element of its structure, its syntax or semantics that did the job (Astington and Jenkins 1999; de Villiers 2005; de Villiers and de Villiers 2009, 2014; de Villiers and Pyers 2002; Hale and Tager-Flusberg 2003; Olson 1989; Tager-Flusberg and Joseph 2005). With time, it was understood, however, that the issue is not so simple and more comprehensive studies proved that virtually all aspects of language facilitate ToM development (Cheung et al. 2004; Milligan et al.

[1] For a description of FBT see, e.g. Perner et al. (1987).

2007; Ruffman et al. 2003; Ruffman et al. 2002), including pragmatics (Furrow et al. 1992; Peskin and Astington 2004; Ruffman et al. 2002).

The issue of the use of language is intimately connected with social factors and so they too have been proven to matter, independently from language (see Devine and Hughes 2016 for a meta-analysis). Correlations have been found between ToM and the number of siblings (Jenkins and Astington 1996; Perner et al. 1994); quality and quantity of interaction with parents (Ruffman et al. 2002); mothers' disciplinary strategies (Shahaeian et al. 2014); mothers' personal epistemologies (Tafreshi and Racine 2016); or, for deaf children, fluency in sign language of the parents (Wellman and Peterson 2013; Woolfe et al. 2002). The general consensus has been that social interaction in general, and social interaction that highlights mental life in particular, facilitates the development of social cognition, including false-belief understanding (Carpendale and Lewis 2006; Galende et al. 2014).

3. There are cultures with folk psychologies that differ from the Western one, and children from such cultures tend to show different developmental timetables and trajectories of socio-cognitive abilities as measured by various tests (Gut and Mirski 2016; Howell 1981; Lebra 1993; Lillard 1998; Mayer and Träuble 2012; Mills 2001; Strijbos and De Bruin 2013; Vinden 1996; Wellman et al. 2011; Wierzbicka 1992).[2]

False belief tests conducted in the above cultures produce significantly different results than those coming from the West (Kallberg-Schroff and Miller 2014). Children from Samoa pass the false belief test at around eight, as opposed to the age of four in the West (Mayer and Träuble 2012). Another Pacific culture—Vanuatu—is similar in this respect (Dixson et al. 2017). Chinese and American children take different trajectories in ToM scale progression (ToM scale is a set of tests designed by Wellman et al. for more fine-grained measurement of ToM than the single FBT can provide) (Wellman et al. 2011). The same progression difference was found for Iranian children (Shahaeian et al. 2011), and a completely novel one in Vanuatu (Dixson et al. 2017). In fact, Dixson et al. (2017) established great differences in the sequence between different social groups within one culture, suggesting that even relatively small differences

[2] A disambiguation of the terms folk psychology and theory of mind is due here. Folk psychology is a common-sense conceptual framework (a theory) that people in a given culture explicitly deploy when explaining, predicting, or manipulating the behavior of other people and higher animals (Churchland 1998, p. 3). Theory of mind, on the other hand, has come to mean in the literature a cognitive system that implicitly guides social cognition and that functions according to the same principle as folk psychology in the West does—namely, a theory of mind. An individual with such a theory-of-mind system will interpret others' behavior in terms of the mind that they "have," which houses various mental states and takes various attitudes towards those states (Gopnik and Wellman 1992).
We know, however, that folk psychology takes various forms and shapes, and that some cultures do not have psychological concepts at all (see Strijbos and De Bruin 2013). However, from the standard meaning of ToM, it should follow that members of such cultures are still guided by a ToM mechanism in their social lives even though they do not explicitly explain social life according to the same principle. The exclusively behavioral folk psychology of the Junin Quecha will be then nothing like the cognitive mechanism they use implicitly (Vinden, 1996), and it will be just a contingent fact that in the West folk psychology has the same or at least largely similar structure to that universal cognitive mechanism. The apparent Western centric nature of that view should raise suspicion, and indeed we will try to show that we could be better off not describing socio-cognitive skills of all cultures in terms of a theory of mind.

in sociocultural context can have great impact on social cognitive development. Kuntoro et al. (2017) drew similar conclusions from their study in Indonesia where they obtained different sequences depending on the city of origin and suggested that the differences in parenting styles between the two cities were responsible. Further, Naito (2014) reports that sixty percent of Japanese children tested did not pass FBT until they were 6 years old. Pakistani children as well showed a lag behind the "standard" Western performance (Nawaz et al. 2014). And a brain imagining study by Kobayashi et al. (2006) demonstrated significant differences in brain areas activated while mindreading in Japanese and American subjects. Finally, and probably most significantly, individuals speaking a newly formed and developing sign language in Nicaragua did not pass the FBT as late as their twenties (Pyers and Senghas 2009).

The traditional two accounts of theory of mind—nativism (Fodor 1992; Leslie eta l. 2004) and rational constructivism (theory) (Goodman et al. 2006; Gopnik and Wellman 1992)—had already formed their respective positions on socio-cognitive development long before most of the empirical findings presented above emerged. When confronted with the reality of these findings, the two approaches had to adjust their models accordingly, in order keep their initial form. For example, nativism had to account for the observed variance in socio-cognitive skills across cultures and social contexts, as well as the influence of language. This has come down mainly to polishing the competence-performance argument (e.g. Helming et al. 2016; Westra and Carruthers 2017). Constructivism, on the other hand, has run against the challenge of explaining the apparent infant mindreading skills. The general strategy for rational constructivists here has been to downplay the importance of the infant experiments, claiming that they do not really require a full-blown, belief-desire theory of mind as such to pass them (Wellman 2014). It is within that climate that the two-systems account has been formulated, which tries to find the middle ground between the two extremes of nativism and rational constructivism (Apperly and Butterfill 2009; Butterfill and Apperly 2013; Low et al. 2016).

Surely, much effort has been made by the three parties to account for the rich empirical data we have available. However, a theory must hold not only on empirical, but also on theoretical grounds. We believe that there are serious theoretical problems with all three accounts. They are untenable because of their foundationalism: they presuppose representational primitives and cannot account for their emergence. Below we describe why this is such a bad thing.

3 Foundationalism of the three dominant theories of social cognition

After Bickhard and Terveen (1995), we define a theory as foundational if it cannot account for the emergence of representational content and therefore must posit a set of representational primitives from which cognitive development starts (see also Thelen and Smith 2002/1996, pp. 28–34). This is untenable because representational emergence is an empirical fact, which should be impossible from a foundationalist perspective.

Foundationalism is a necessary consequence of theories that view mental representation to be encodings—symbols with semantic content that refer to the outside reality.

There is just no way for such representationality to emerge from non-representational phenomena. All three traditional ToM accounts share that view of representation and hence they too are foundationalist, regardless of their particular differences in accounting for the empirical data. We demonstrate this in the next section. Our criticism is deeply indebted to Bickhard and Terveen (1995), who offer a detailed criticism of foundationalism in cognitive sciences.

3.1 The problem of emergence and the necessity for foundational concepts

Natural cognition is the ability to acquire information from the environment, retain it, and use it for the purpose of adaptive behavior. Mental representation is argued to be the central process making this possible. We should ask, then, what it means for mental representation to be in service of information acquisition, retention and behavior guidance. At the most general level, it means that the organism understands something about the represented reality: representation should give a clue to the organism about what it can *expect* to happen, what it can *do*, and what it *should do* considering its goals and current states.

Traditionally, representation is argued to achieve the above in virtue of its correspondence to the represented reality and systematic relationship to other mental representations (Cain 2013; Fodor 1975, 1983; Pylyshyn 1984; Smith and Medin 1981). This has classically been viewed as a causal, informational, and ruleful relationship between the representation, the represented, and other representations of the representational system. To capture that, mental representations have been assumed to have semantic properties much like logical terms, propositions and propositional attitudes do; they refer to reality via the meaning encoded in them and to each other via their syntactic properties. Consequently, much of cognitive psychology today views cognition in terms of manipulation of semantic symbols in this sense. Environmental information is said to enter the system through the senses, after which it is transduced into a symbolic or representational format that is independent from the specific sense modality through which the information was acquired. That is when sensory representations become concepts. Once in the amodal format, information is processed in a way similar to logical inferences. Finally, results of this computation are transduced back into the embodied format of motor directions.

There is a problem, however, in accounting for the emergence of the amodal semantic content—concepts must be already in place for the sensory information to be transduced into them and back into motor information (Bickhard and Terveen 1995). In fact, emergence of semantic content is impossible; there is no way for the organism with such a code to actually know what it is about (for the criticism see Bickhard 2009b; Bickhard and Terveen 1995), and for that reason any model built around semantic representation is forced to be foundational—to necessarily assume a set of conceptual primitives from which development can start. Below, we point out how the traditional ToM approaches share this problem.

3.2 Nativism

Researchers with nativist views generally follow the standard modularist model of cognition (Carruthers 2013, 2015, 2016; Helming et al. 2016, 2014; Leslie 1994; Scholl and Leslie 1999). As with similar accounts of other cognitive abilities (e.g. Lightfoot 1989; Pinker 2014/1994; Wynn 1992), the story here goes as follows: There is an innately specified cognitive mechanism, or module, dedicated to a specific domain, in our case—to mindreading. This mechanism is independent from and insensitive to virtually any extra-organismic factors and develops according to a biologically predetermined schedule. The innate, encoded information that it contains consists of basic inborn concepts (e.g. BELIEF, DESIRE, SEE, and PRETENSE) and heuristics (e.g. "seeing leads to believing") that enable the child to pick out relevant stimuli, cognize it, and draw basic conclusions (in an unconscious modular way, that is). These are then fed in some form to the central system, adding an aspect of another's agency to the child's perceived reality.

Nativists take this to be the cognitive process that underwrites 15-month-olds' performance on the spontaneous–response FBT; infants pass it because they have this basic inborn theory which makes them expect the false-belief congruent scenario. As Westra and Carruthers (2017) offer, this innate theory is open for learning; it is claimed that it gets enriched with more complex concepts throughout development, or that its harmonization with the rest of the cognitive system can improve, domain-general processes putting the mindreading module to work for their purposes (Carruthers 2015).

Accordingly, any differences in performance on the elicited-response FBT and related tests across populations—e.g. different cultures, parenting practices, or linguistic inputs—are explained away by factors other than an actual lack of mentalistic understanding of the mind (this is most clearly argued in Westra and Carruthers 2017, but see also Helming et al. 2014, 2016). Nativists take two directions here. One is the claim that the tasks in question make demands for more complex concepts and heuristics than the basic ones (culturally embedded ones not yet acquired). The other is the recourse to the competence-performance chasm. The latter path is naturally necessary for FBTs. Children are said to understand false beliefs innately and the varying performance is due to (a) misconstruals of implicatures of the test questions, (b) lack of adequate vocabulary in the language they are growing up with, or (c) undeveloped executive functions or general-processing resources.

Nativism is thus openly foundationalist, and seeks a solution to the impossibility of emergence in the claim that the representational primitives are innate. This does not help, however, as the idea of inborn semantic content cannot be defended either. We discuss this in Sect. 4.

3.3 Rational constructivism

The alternative, rationalist constructivist idea has been that children construct a theory of mind much like a scientist would (Gopnik et al. 1999; Gopnik and Wellman 1992). This faced a challenge in light of the spontaneous–response FBT results with infants.

In order to explain the gap between infant and preschooler mindreading, and not to commit to innate belief understanding, constructivists were forced to show that infant mindreading could be explained by simpler concepts than those of a belief-desire theory of mind.[3]

Wellman (2014) argues that infant data can be explained with only a desire-awareness conceptual framework, which with time is built on and becomes a proper theory of mind with belief understanding. The way the original concepts are modified and enriched is modeled with the use of Bayesian hierarchical networks, and the child's conceptual development is viewed as a theory revision process, the child being a "little scientist." Thus, in contrast to the nativist view, constructivism of this kind views ToM development as utilizing domain-general resources, and conceives of ToM as a somewhat real theory in the mind of the child, not a modular mechanism operating according to the same principle as a theory of mind. It does, however, start with a set of foundational mental representations (Wellman 2014, p. 197), and necessarily so. For a theory-based development, the initial states must be concepts understood as encodings; theory by its nature just has to start with initial concepts that enable hypothesis formation and further theoretical change. It is fairly unquestionable that any account has to start with something, to take something for granted. Rational constructivism, however, forces us to assume that these initial states are conceptual in nature, and this is clearly a case of foundationalism.

3.4 Two systems

The two-systems view finds a middle ground in between nativism and constructivism. Although there have been a number of different proposals that advance two systems (e.g. De Bruin and Newen 2012), Apperly and Butterfill's account is most usually associated with the term (Apperly 2012a, b; Apperly and Butterfill 2009; Low et al. 2016).

The basic assumption of the two-system view is that children pass the spontaneous–response FBT because they are in possession of mindreading system 1, which is a limited foundational theory of mind: a belief-tracking mechanism (Apperly 2012a; Apperly and Butterfill 2009; Butterfill and Apperly 2013).

Imputing spontaneous–response FBT results to the workings of system 1, two-system proponents claim that passing the elicited-response FBT requires a much more effortful and explicit way of thinking about other minds—what they call system 2—that children before preschool cannot really use. Although Apperly and Butterfill do not really address the development of system 2, it would make sense that it is constructed or somehow acquired just as other explicit ways of thinking about reality, and so its development might be influenced by environmental factors, which would explain the cross-cultural data we have presented before.

[3] It is important to note that constructivism is a wide term, and, in fact, the action-based approach we advocate in this paper is a version of constructivism as well. Since we are discussing traditional accounts at this point, however, our focus and criticism falls on the traditional rationalist or theory theory strand of constructivism advocated by Wellman and Gopnik (Gopnik and Wellman 1992; Gopnik and Wellman 2012; Wellman 2014).

Theoretically speaking, however, two-systems does not offer a way out of foundationalism. Although Apperly and Butterfill are much more subtle than the proponents of the other accounts in their distinction between a theory-of-mind ability (an ability to behave as if one had a theory of mind) and theory-of-mind cognition (using a theory of mind as such), they still view their system 1 in terms of the latter, albeit not "full-blown." And with it, necessarily come its encoded representations:

> We do not aim to argue that someone could track beliefs, true and false, without any theory of mind cognition at all. Our concern is rather with the construction of a minimal form of theory of mind cognition. As we shall explain, minimal theory of mind does involve representing belief-like states, but it does not involve representing beliefs or other propositional attitudes as such. (Butterfill and Apperly 2013, p. 3).

Further, it does not change matters much that they view their minimal theory ascribed to infants as only a construct at the computational level of explanation (cf. Marr 1982/2010). The computational level explanation still imposes significant constraints on possible implementations. Apperly and Butterfill's framing of the computational problem in terms of a minimal theory of mind still poses a foundationalist problem: The computational-level theory has to be implemented in a way that merits being called "a theory"—that is, there have to be implementational equivalents of computational-level processes that relate to each other in the prescribed, theory-like way (compare to the discussion on tacit knowledge in Davies and Stone 2001; Fenici's 2013, application of Davies and Stone's ideas).

The computational problem they describe still consists in ascribing states to observed agents. Thus, system 1, though minimalistic, still presupposes concepts of object and agent and registration, whose contents are similarly left unexplained developmentally. As far as system 2 is concerned, here they run into all the problems that rational constructivism does—they openly draw an analogy to Fodor's central system (Apperly and Butterfill 2009, p. 956), where explicit theories reside, when explaining why system-2 theory of mind should be effortful but flexible.

Notably, it can be argued that the two systems account does not aspire to explain the development of the representational states in question, and thus the charge of foundationalism does not apply to it. Still, the problem generally has to be solved to give an exhaustive account of socio-cognitive development, and there does not seem to be much potential in the two-systems account to do so, as it frames the problem in cognitivist terms.

The general insight of the two-systems theory, however, is consistent with the interpretation offered by our action-based account. Following the action-based principle, we too arrive at the closely similar conclusion that there is an important chasm between processes underlying competent social interaction, and explicit theorizing about other's mental life. How we reach that interpretation is however importantly different, as later parts of the paper will show.

4 Emptiness of the concept of innateness

As we demonstrated above, all three accounts are foundational, which renders them theoretically untenable. One argumentational move that is often employed by foundational accounts is to defend foundationalism with a recourse to nativism. This is explicit in the nativistic accounts (e.g. Carruthers 2013, p. 151), but also a potential response of the other two frameworks. Therefore, below we point out why nativism is untenable in its own right.

As Racine (2013) argues, core or foundational knowledge approaches tend to use a neo-Darwinian adaptionist view on innateness, claiming that the inborn knowledge and skills present in infancy were an object of natural selection in phylogeny due to their evolutionary advantage, and hence are *coded* in genes and necessarily present in every individual. Rather than being solved, the foundational weight is thus moved onto biology. However, the move is unwarranted as developmental biology speaks against phenotypic traits as complex as concepts being formed prenatally and irrespective of experience.

Great revisions are afoot in modern biological sciences as some consider the twenty-first century to be the century of biology (Venter and Cohen 2004). One of the central issues in this revolutionary climate is precisely that of evolutionary mechanisms and viable notions of innateness. Following works of such researchers as Lewontin and Gould, modern biology is much more cautious with adaptionist stories of traits and the idea of them developing "innately" in ontogeny (Gould and Lewontin 1979; Lewontin 2001; Oyama 1985/2000). Psychology, however, seems much slower to catch on to this trend (cf. Racine 2013), as we see evidenced by the foundationalist accounts of cognitive development.

As biological research demonstrates, development is a multiply contingent process (Elman 1996; Gould and Lewontin 1979; Gould and Vrba 1982; Mameli and Bateson 2011; Oyama 1985/2000; Pigliucci and Müller 2010). A number of psychologists urge researchers to consider this in cognitive development too (Carpendale et al. 2013, p. 130; Carpendale and Wereha 2013, p. 208; Lewis et al. 2013, pp. 159–160; Lewkowicz 2011; Spencer et al. 2009). They point out that there are multiple elements whose interaction leads to the development of biological and cognitive forms, and hence any talk about "innate," meaning encoded in genes, contorts the way in which genes matter for development. The "interactivist lesson" taken from the discussions in biology is that genes have their developmental significance only in the context of other intra-organismic as well as extra-organismic interactants. In other words, they have their "information" about a particular form only inasmuch as we keep other causes constant, which is hardly the case in nature. This interactive nature of development renders any talk about "genetically specified" innateness meaningless. We would be making just as much sense talking about innateness being "environmentally specified" since for genes to have their particular causal powers there needs to be a particular environmental context (Carpendale and Wereha 2013, p. 208).[4]

[4] One can argue that what is actually meant here is psychological innateness, which is just a *methodological* foundationalism about certain traits that fall outside the scope of psychology (see, e.g., Samuels 2002). That is, the concept of, say, belief could be taken as a primitive in psychology because of the fact that the intricacies

We argue that the nativist ToM approaches assume the idea of innateness that no longer fits with current research in biology and therefore construct their theories in a theoretical vacuum. Let us have a look at this excerpt from Carruthers (2013):

> The infant-mindreading hypothesis, in contrast, postulates an innately channeled body of core knowledge, or an innately structured processing mechanism (or both), with an internal structure that approximates a simple theory of mind. The explanatory burden, then, is an evolutionary one: it needs to be shown that there were sufficient adaptive pressures among our ancestors for such a mechanism to evolve. There is now an extensive body of work suggesting that this is indeed the case. The gains provided by such a mechanism might derive from enabling so-called 'Machiavellian intelligence' (Byrne and Whiten 1988, 1997), from facilitating larger group sizes (Dunbar 1998), from enabling distinctively human forms of cooperation and collaboration (Richerson and Boyd 2005; Hrdy 2009), or from any combination of these. (Carruthers 2013, p. 151).

According to Carruthers, a main challenge for the nativist explanation is supposed to be telling an adaptionist story. However, this contributes little to developmental models because phylogenetic adaptations *are not* preformed phenotypes that are necessarily expressed in ontogeny.[5] Neo-Darwinian adaptionism is a meta-theory conceived to talk about phylogeny exclusively: "The neo-Darwinian framework is at root, by definition, a nondevelopmental framework" (Racine 2013, p. 144). We may talk about innate features in phylogenetic analyses where the term is used to mean "reliably present in the species in a given environment"; these analyses assume developmental contingencies to be constant and talk about changes in population over phylogenetic time. When we are interested in ontogeny, however, we are trying to figure out precisely that which is excluded from the neo-Darwinian adaptionist framework—contingencies of development and how phylogenetic heritage interacts with the actual context of growth. The fact that some trait evolved through adaptation does not mean that it is innate in the sense that most nativist theorist seem to assume—that it is preprogrammed and necessarily present regardless of environment (Oyama 1985/2000, p. 25).

Adaptations happen in an environmental context and certain aspects of that context are usually necessary for them to develop in ontogeny. What is then "innate" is not an intraorganismic encoding that is the problem of the evolutionary biologist to explain, but an integrated organism-environment stability that any developmental account must tell the story of.

Footnote 4 continued
of its development are fully explainable at the level of biology. There would be no need to explain them in psychology same as there is no need to explain the development of limbs or internal organs. However, the problem is that the psychological primitives assumed in cognitivism *are not* explainable at the level of biology; in fact, as we argue here, there are serious reasons against mental representation developing without processes falling within the ambit of psychology. This fact tends to just be ignored in nativistic accounts. As Lewis et al. (2013, pp. 159–160) argue, "we don't design a perpetual motion machine and then say that building it is an engineer's problem. We need to say what is actually meant by "innate," other than saying 'It's not my department, ask a biologist'."

[5] Moreover, biologists warn against overhasty adaptionist stories that one can tell; although useful in the current environment, certain features may not have been selected for but emerged as a result of what is called, after Gould (1991), *exaptation*.

Accordingly, even if a cross-cultural universality is established in infant performance on the spontaneous–response FBT, this does not entail that the necessary cognitive skill develops innately. The universality is most likely due to similarities of experience across these cultures, not to a genetically or internally specified module. This means that not only more cross-cultural spontaneous–response FBT studies are needed, but also inquiries into the nature of the contexts of growth in the cultures studied, which would make it possible to identify potential similarities and differences that can modulate the development of the skills. Only once these potential environmental modulators have been excluded as a partial cause of socio-cognitive skills could we advance any nativist (i.e. developmentally internalist) claims.[6]

In sum, empirical and theoretical considerations about development speak in favor of the view that evolutionary endowment interacts with other factors in ontogeny and leads to social competence, rather than providing a preformed ability or representation. Consequently, the nativist ToM accounts are stuck between their inability to account for the emergence of representation in ontogeny and the implausibility of representation forming in phylogeny.

Below we present the action-based framework that solves the problem of foundationalism and is consistent with developmental science. Finally, we offer a sketch of an action-based account of the three groups of empirical data.

5 Solutions offered by an action-based perspective

In recent years, we have been witnessing a pragmatic turn in cognitive science (Engel et al. 2015) as various action-oriented views are proposed to redress the flaws of classic symbol-manipulating models. Although contemporary action-based approaches to cognitive development are still in the works (Pezzulo et al. 2015, p. 49), the central importance of action has been recognized by a number of theories, both older and more recent ones. To name a few: Piagetian approaches (Allen and Bickhard 2013; Bickhard 2009a, b; Carpendale and Lewis 2004, 2006; Newcombe 2011), Vygotskian approaches (Nelson 2007), dynamic systems (Thelen and Smith 2002/1996), grounded cognition (Barsalou 2008), radical enactivism (Hutto and Myin 2013, 2017), or the Predictive Processing Theory (Clark 2016).

Although there is much work to be done before we arrive at a comprehensive action-based account of cognitive development, the action-based principle has a lot of potential to create a much-needed unifying framework for development. Here we are interested in what the framework can offer to the research on social cognitive development. To explore this, we briefly sketch the action-based principle and demonstrate how it deals with the problems of foundationalism and nativism. Then, we provide a provisional action-based interpretation of the three main groups of data about social

[6] Meristo et al.'s (2016) findings are relevant here (see also Meristo et al. 2012). In their study, deaf infants did not pass the spontaneous-response FBT, as opposed to hearing infants. This suggests that even this early social competence is not "innately" given, but rather develops through interaction with the environment. Meristo et al.'s (2016) own interpretation focuses on necessary family interaction, in keeping with the findings we cited earlier (see Sect. 2).

cognition, and stress the interpretation's fundamental difference from the classic cognitivist ones.

In our sketch, we draw on three frameworks: Bickhard's *interactivism* (Bickhard 2009a), Carpendale and Lewis' (C&L) *social-experiential approach* (Carpendale and Lewis 2006), and Nelson's *Community of Minds* (Nelson 2007). These frameworks, although largely underrepresented, have a lot to offer to the current debates in social cognition, especially in reference to the problems we have discussed in this paper. We do refer to other compatible and relevant theories in passing, but do not wish to present an exhaustive review of this sort.

5.1 The action-based principle

Here is what follows from the criticism of the ToM approaches we offered above, which we contrast with what the action-based principle claims: (1) For the criticized theories, every action or cognitive skill is underwritten by disembodied representational competence of sorts; for action-based models, action can precede representational mental content.[7] (2) For the criticized theories, the development of social competence *must* start with an inborn base of amodal representation; for action-based models, it does not have to—representation can emerge from non-representational phenomena.

It is instrumental to stress at this point that the action-based representation (which we take from interactivism) and the standard idea of representation as amodal encoding differ fundamentally.[8] First, what we are interested in when modeling representation is not solving the metaphysical problem of reference (see, e.g. Quine 1960/2013, pp. 23–72), but only proposing such an idea of representation which is a viable way in which real organisms can represent reality. An action-based representation does not represent on the basis of reference or correspondence; it is not a disembodied symbol with a semantic stand-in for what is being represented. It does, however, have the necessary properties of representation—intentionality and truth value. And most importantly, it has them in virtue of processes which are consistent with developmental reality and which allow for representation to emerge in ontogeny (and phylogeny) from non-representational phenomena. How it does so, and how it achieves intentionality and truth value should become clear in our exposition of the action-based principle below. This is drawn from interactivism (Bickhard 2007, 2009a, b, 2010; Bickhard and Terveen 1995).

An internal state S is a detection of an external state S*. The organism does not know that, but merely experiences the internal state as "this state." Being in internal state S, the organism undertakes action A (from among others; let us assume that for newborns actions can be random at first for the sake of the illustration), which

[7] This is made explicit in interactivism (Bickhard 2009a, b), but see, for example, "use without meaning" in Nelson (2007).

[8] It needs to be noted at this point that many other action-based approaches are anti-representational (e.g. Hutto and Myin 2013, 2017; Thelen and Smith 2002/1996; van Gelder and Smylie 1995). Interactivism, however, provides what is to our minds a convincing account of why we should keep the concept under the revised meaning. This theoretical difference notwithstanding, in virtue of the action-based principle, we believe that the anti-representational alternatives would yield similar empirical claims as we sketch here with the use of interactivism and the other two frameworks.

results in the external state S* changing to Y*. The organism's physical organization is such that Y* evokes another internal state, state Y. This way, on the basis of non-representational detective properties, the organism can create an action-internal state contingency pattern that while in state S, action A leads to state Y. This provides the germ for normativity—the organism will now (implicitly) know that state S is not only just "this state" but also such a state that can lead to state Y via action A. Thus, the organism functionally predicates something about the current situation; and it does so in virtue of the action-internal state contingency it has the knowledge of, without the need to refer to the outside world at all. Moreover, the predication can be false or true, and the truth value can be potentially known to the organism—all it takes is to engage in action A to find out whether it is possible to go to state Y from state S. If it was not, then the previous situation was not the situation that should have produced state S (no external state S*) or state S needs some other states co-occurring in order to afford going to Y, which were only accidentally present in the previous interactions that went from S to Y via action A. The organism can accordingly update its functional predications after failing to achieve the expected result of its action.

Now, there may be processes in more complex organisms whose main function is to probe the external reality in the way presented above. They can serve to keep track of what interactions are possible in the given situation so that the organism can be a competent agent that can choose from an array of affordable actions in light of its current goals. Bickhard refers to these kinds of representational phenomena as *apperception*. They are much more like the classic idea of representation as their function is mainly not to directly indicate the possibility of a given action, but only to predicate something about the current situation. This predication can then be the basis for many other interactions and constitutes what Bickhard calls the *situation knowledge*. Thus, apperceptive processes *have the function of representation* for other interactive processes that rely on them in order to guide adaptive behavior. In fact, any representation of a possible interaction can serve two functions—to cue the organism to engage in the interaction, or to provide information about the situation for another interaction representation (something like knowing that you can order a taxi anytime at a party makes you entertain staying after the last bus home leaves). Moreover, there are most likely many levels of recursive interactive systems within the architecture of the human mind, in which one system interacts with interaction potentials (representations) of a lower one, enabling explicit thought about its properties (Bickhard 1998; Campbell and Bickhard 1986). However, it is the function of representation for the organism's processes that makes it representation, not some amodal, symbolic format it is coded in.

Importantly, the action-based representation offered by interactivism is inherently embodied and situated. The content is constructed through interaction and is grounded in the modalities of the experiences. (Ap)Perceptual processes constituted by sensorimotor contingencies (SMCs) may be grounded in one particular modality (O'Regan and Noë 2001), but the situation knowledge they create, and on which higher-order interactive representations rely, will span all the modalities. Representing the car, for instance, will base off SMCs grounded in past explorations of how cars look, smell, feel, what sound they produce, and for some maybe, what they taste like. The interactive representation of car will consist in connecting the expectations of all these

modalities under a common contingency—if I hear a car, visual contingencies associated with it activate too and I expect to see it when I turn my head. The central point is that the content of action-based representation is modal by definition as it is past experience that constitutes it.

Consider a simple example (this is just to make a point, not an empirical claim). The infant experiences a state of hunger and starts crying (crying action could occur randomly, but it is plausible that such a simple state-action coupling could form prenatally; note, however, that the infant does not know anything about why she acts this way). Crying has its impact on the environment such that the mother comes and starts feeding the baby. As a result, the state of hunger changes to the state of satisfaction. Thus, the infant functionally and implicitly comes to know that when hungry, crying leads to satisfaction, and she has some (again, functional and implicit) idea what the internal state of hunger "means". This knowledge—of hunger-crying-satisfaction contingency—does not have to be innate as the information about it is reliably present in the environment of growth; hence, evolutionary selection was more likely to predispose the child to quickly learn it in the above way rather than prewire it whole.

If the mother is not in the room, however, then crying will not have its effect.[9] Making sure that the mother is in the room will therefore be an apperceptual process in service of hunger satisfaction via crying. Perception is an apperceptual process par excellence, and we have an action-based account of perception nicely worked out by O'Regan and Noë (2001).[10] This way, through apperceptual processes, the child keeps track of whether the given situation really is a situation in which crying will lead to feeding and satisfaction of hunger.[11]

The way that language changes the interactive context needs to be noted as we refer to it in our account of the three groups of empirical findings. Linguistic interaction will build on non-linguistic interaction patterns and words will come to represent interaction possibilities they have been used in. Consider a game in which a linguistic utterance is a part of—a simple naming game that mothers play with children many times, where when presented with a toy, one has to say its name. Grounded in such an interactive pattern, the word Zebra, for instance, can come to represent the toy as it was used in the game upon the presentation of the toy Zebra (feedback for failures in naming it so could have been provided in the form of undesirable interaction on the mother's part—negative facial expressions or continuation to hold the toy instead of progressing to the next one). Linguistic units grounded in such a way can be naturally uttered in any situation, and the child will learn this when her mother uses the same

[9] Let us ignore the possibility of having a baby monitor in the room for the sake of the example.

[10] O'Regan and Noë (2001) have provided a well-rounded account of how this basic detection and historicity combined with action can explain development of perception. The changes in sensory states effected by various actions (e.g. movement of the eyeball for vision) lead to the establishment of sensorimotor contingencies (SMCs) that constitute perception: What we perceive (also phenomenally) is an implicit memory of perceptual states possible to be achieved in the current perceptual state if certain actions are undertaken. From an action-based perspective, even associations acquired from an observer viewpoint are thus actively constructed as they are the result of ongoing prediction, validation or falsification, and revision of expected sensory states.

[11] Why the child strives to satisfy her hunger in the first place is explained by viewing living organisms as self-maintenant far-from-equilibrium systems, see, e.g. Bickhard (2010); for the present purposes, however, it suffices to say that the organism is organized in such a way that ensures its ongoing existence.

expression in a different context, which evokes the interaction potentials in the child's mind grounded in the past use of that word. Something similar will take place when the child herself uses it in a different context and observes its impact on the external, social world. Importantly, further contexts of use can be themselves linguistic, which is possibly how abstract meanings emerge.[12]

There is much more to be said about language in an action-based framework (Bickhard 2007). However, for the present purposes, we want to point out that some socio-cognitive abilities could require language to develop, while some would not. There are some social competences which will be greatly improved by associating them with linguistic interaction, some that are embedded in linguistic interaction entirely, and there are such that do not gain much from it. For instance, mere physical sequential social interaction of changing the diaper would not gain anything by adding linguistic components to it, whereas going for a walk would (linguistic structuring of activities outdoors makes them both safer and more interesting for the child). And it goes without saying that kinds of interactive competences that are entirely embedded in language—such as being able to hold a conversation—would need previous linguistic experience in order to exist at all.

It should be clear after this exposition that an action-based perspective solves the problem of foundationalism and is consistent with the multiply contingent nature of developmental phenomena. In fact, many of the aspects of the model mentioned above—such as emergence of representation from non-representational phenomena, naturalized normativity, system-detectable error, or the possibility of multiple, interrelated but qualitatively different representational processes within the organism—are simply absent in the framework upon which the traditional ToM accounts are based. As such, an action-based perspective offers a much more comprehensive alternative to modeling socio-cognitive development that can replace or potentially complement the standard models.

5.2 Constraints on experience in an action-based framework

With the above model of cognitive development, the kind of interactive experience the child engages in will determine the content of its cognitive structure. It is possible that some internal state-action contingencies are already present in newborns (think of all kinds of reflexes), but it does not seem likely that they come with prewired complex interaction representations of theory of mind. A more viable thesis is that the macroarchitecture of the nervous system predisposes the child to certain kinds of experience, and it is these experiences that form its microstructure, furnishing the mind with representational contents (understood pragmatically, not semantically). Natural selection has happened in the environment where certain elements were reliably present and so whatever developmental process it has selected will implicitly presuppose their presence in ontogeny. In other words, the phylogenetic heritage determines what can

[12] Note also that the acquisition of the syntactic structure of a language should be acquired in just that way—through interactions where words have their function only in the "proper" syntactic position (see Bickhard 2007, p.183).

be possibly experienced by the child, but it is the context of growth and experience therein that determine what *will* be experienced (cf. Nelson 2007, p. 249).

The kind of interactive experiences available to the child will naturally differ depending on the time and place of development. Nelson's (2007, p. 19) model of constraints on cognitive development captures this fact nicely. She identifies six kinds of constraints: evolved, embodied, ecological, socially embedded, encultured, and that of past experience.[13] What is experienced at a given time in development is jointly determined by the constraints. The view on development that we get here is therefore necessarily scaffolded—starting from the interactions available to an infant, she actively "acts her way up," establishing first basic action-based representations (e.g. sensorimotor contingencies), and then using them to engage in new interactive experiences and establish further, more complex, representations (e.g. social competence).[14] The basic representations will be fairly invariant across environmental contexts as they will rely on largely universal patterns of physical interaction with the environment and caretakers; higher representations will be grounded in more complex social and cultural interaction, which will naturally differ greatly across social and cultural contexts.

5.3 Socio-cognitive development in an action-based framework

Coming back to the problem at hand, two constraints on experience are especially relevant for the emergence of socio-cognitive mental structure—non-linguistic social interaction and linguistic social interaction. Below we show why,

Carpendale and Lewis (2004, 2006) stress the role triadic interaction with parents and objects has for cognitive development. Naturally, the kind of interactive experience the child gets while interacting with other people will establish socially embedded representations of interactive potentials: The child will have developed certain expectations of how people behave in a range of situations and how they react to the child's own actions. Such largely behavioral interaction competence will not, however, involve abstract concepts of minds or beliefs. The action-based representations established will be derived from purely physical interaction with other people, not understanding their minds. As argued by others (e.g. Fenici 2012, 2015; Gallagher 2008), such embodied interactive competence can readily explain early social cognition and spontaneous–response test results.

Things change when language enters the developmental system. Conversational situations constitute a new kind of interactive context that enables representation of unobservable minds. Carpendale and Lewis claim that language is acquired by learning patterns of interactions for which "it is appropriate to use a particular term, for example, mental or emotional, or dealing with pain, and so forth" (Carpendale and

[13] The past experience constraint is exactly what we described as the knowledge of interactive potentials. Current knowledge of interactive contingencies naturally constraints what can be experienced and what further representations can be acquired.

[14] Note that such a view conveniently allows for the possibility of other mental processes—e.g. affective—being implicated in social cognition (see Carpendale and Lewis 2006). This contrasts with the almost exclusively cognitive focus of the theory-of-mind research; it is not clear how the operations of mindreading mechanisms or proto-scientific theories could be influenced by emotions and in fact, there has been little research in that direction.

Lewis 2004, p. 88). Linguistic interaction is part of the external environment and as such provides interactive potentials embedded in language, otherwise unavailable. This fact—that language enables later social cognition based on abstract notions of mind and belief—has also been argued by others on various grounds (e.g. Fenici 2012; Gallagher and Hutto 2008; Hutto 2008; Nelson 2005, 2007). An action-based principle could provide a more detailed model of why this is so. The bottom line is that linguistic interaction potentials can go a long way in explaining how children pass the standard FBT, without the need to posit theoretical knowledge (cf. Gopnik and Wellman 1992).[15]

5.4 Empirical data within an action-based framework

Now we can turn to the difference an action-based approach makes in the interpretation of the groups of findings we reviewed at the beginning of the paper. As already discussed, social interaction is necessary for the emergence of any representation that pertains to other people as such representations originate in past social interactions. Language is necessary for representations that are embedded in *linguistic* social interaction, and facilitates those that benefit from linguistic input. For the development of folk psychology both linguistic and social interaction are naturally necessary (see Fenici 2017; Fenici and Garofoli 2017; Hutto 2008).

1. Infants' performance in spontaneous–response FBT can be accordingly explained as a certain point in the development of interactive competence that 15-month-olds are at. Over the span of their lives so far, the infants have had enough interactive experience to represent and expect a false-belief-congruent scenario in the spontaneous–response FBT. From an action-based view, this is however only functional competence, not underwritten by abstract concepts. The representational processes involved are interactive potentials and expectations that follow from them, grounded in past interactions. More specifically, perhaps the infant's expectancy is based on the chronic helpfulness in such situations that children of that age exhibit (Warneken and Tomasello 2007), combined with their experience of situations where previously absent adults interact with the kid and the toys in the room in a way that betrays ignorance about the toys' location.[16] To wit, the expectation that is violated in spontaneous–response FBT has its source in the infant's

[15] Incidentally, language possibly enables abstract thought in general as it amplifies *every* situation with its interactive potentials. As Carpendale and Lewis argue, if I am a user of language, every situation has the potential of me saying anything (see Chapman 1999, p. 34 who C&L cite). This ipso facto affects the potentials of the situation, which now affords linguistic interaction. Note that words, like any other interaction, are grounded in the past experience of their use (i.e. their meaning is embodied on the basis of internal states experienced as a result of conversations in the past). That is, words' meaning is the anticipation of an internal state grounded in their past use. Once established in the brain, they allow me to entertain that anticipation in any situation (as words can always potentially be said), and can effectively represent any situation where language was involved. Language in this view would allow the specifically human kind of thinking that can branch out to the farthest possible scenarios. Abstract concepts would be then those kinds of interactive potentials that are grounded in purely conversational contexts of use (see Nelson 2007, pp. 151–152).

[16] The experience of adults leaving the room and coming back, unfamiliar with the changes in the set-up of toys in the room does not seem to be an uncommon situation in a 15-month-old's life.

representation of the current interaction, not mental states. It is possible that the surprise consists in the falsification of the interaction representation where the child helps the adult and the adult's established and expected role is not knowing where the objects are when she or he comes back into the room.

This is, of course, just a provisional projection; a more thorough analysis is needed, and more empirical tests done in the action-based spirit would illuminate the issue as well. What is important for us now, however, is that from an action-based perspective, the psychological phenomena behind the performance are *not* inborn or foundational, but constructed by past interactions, and the mental representation involved pertains to the physical aspect of the interaction only (cf. Banovsky 2016). There is no knowledge about unobservable minds involved. Infant social cognition is in this view competent social interaction *based on previous non-linguistic interaction*. This is generally consistent with minimalist accounts of early social cognition and some of the theoretical solutions offered there (Fenici 2015; Heyes 2014; Hutto 2015; Perner and Ruffman 2005; Ruffman 2014; Ruffman and Taumoepeau 2014). It is not consistent, however, with the two-systems account and their not minimalist enough account of early social competence as it concedes that system-1 agents still do understand *something* about mental life.

2. It should be clear by now that the significance of linguistic and social factors for social cognitive development is completely reevaluated from an action-based perspective. First, to have any expectations about social situations, infants need to have had relevant interactive experience that has led to the capacity for the formation of the respective situation knowledge—other people are part of the interaction potentials that the child represents in any situation that is social. How other people have behaved in the past and how they reacted to the child's actions determines the kind of representation the child will have of any particular social situation. Without language, however, this kind of representation builds on purely physical interaction. Language introduces its unique interactive possibilities that are grounded in past linguistic interactions and make it possible to represent what is not physically there. Seeing Sally come back to the room in the FBT (Baron-Cohen et al. 1985) will induce interactive potentials of not only the observable reality, but also of linguistic interaction, rooted in past experience that involved talk about minds and intentions. It is only through language, then, that one can talk about the formation of abstract concepts; minds, beliefs, desires and other mentalistic concepts are developed through linguistic interactions that the child takes part in. Once there has been a sufficient amount of such linguistic experiences, it becomes possible to think about other minds too regardless of the current situation, as language imbues every situation with its interactive potential (cf. Clark 1997, pp. 193–218; Dennett 1996, pp. 147–152).

Consequently, the social-linguistic factors present in previous experience will naturally influence empirical results, as these factors are what largely constitutes the child's representational abilities. In other words, more sophisticated social cognition, e.g. such that involves understanding opacity of terms, or one that involves giving justification of one's answer, will build on previous *linguistic* interaction. This is in line with the fact that all aspects of language matter for social cognition (Milligan et al.

2007); this empirical observation makes sense in the current framework because language is not a computational tool that provides some special format or syntactic tool for symbol manipulation, but rather an interactive system that permeates the child's cognitive structure (Bickhard 2007).

3. Anthropological research informs us that linguistic and social interaction differs greatly across cultures and therefore affords often starkly different interactive experience. It is true that we may obtain similar results from the spontaneous–response FBT with infants across different cultures—the kind of behavioral interaction that provides content for the expectation tested there can be fairly universal. As Carpendale and Lewis have it, "children […] may achieve comparable levels of development at similar ages because of commonalities in their experience" (Carpendale and Lewis 2004, p. 85). Relevant linguistic interactions, however, are evidently extremely different in many cultures and therefore language-embedded representation of other minds will be such too. The same applies to the narrow context of family and friends—they too afford varying interactive experiences and influence cognitive development, which is evident in the empirical data we presented at the beginning of the paper. The way other people are talked about, the way that social interaction is narrated, and the role that linguistic behavior plays in social interactions determine the kind of abstract representation of mental life that members of a given culture or family construct. We observe, for example, Japanese children failing the verbal FBT, not because of the fact that they predict the searching to be incongruent with the false belief, but rather because their *justification*, which is necessary to pass some versions of the test, does not refer to minds and beliefs, but to social relations (Naito 2014). This is not a surprise from an action-based perspective as Japanese children have had different linguistic interactions in their past than children from the West, based on which they have formed different interactive potentials (representations). It is then entirely natural from an action-based perspective that we tend to find the greatest differences in social cognition in cultures that differ from the West in both social (family relations, social conventions, philosophical traditions etc.) and linguistic (syntax, semantics, and pragmatics) respects. Representational underpinnings of folk psychology will genuinely differ across cultures inasmuch as social and linguistic experience differs in them in relevant aspects (cf. Fenici 2017).

It becomes clear that the view is generally consistent with other views that proclaim the embodied nature of early social cognition, and linguistic nature of later mindreading (e.g. Andrews 2012; Fenici 2012; Fenici and Garofoli 2017; Gallagher and Hutto 2008; Newen 2015; Zawidzki 2013). However, the exact views on the nature of cognitive processes involved espoused by these approaches could differ greatly from the action-based ontogeny we have presented.

6 Conclusion

We have shown that nativism, rational constructivism, and two-systems theory offer unsatisfactory explanations of socio-cognitive development for two related reason-

s—their foundationalism and nativism. All three theories are foundational in virtue of their encoding-based view on mental representation, which precludes representational emergence a priori. Appeals to innateness do not offer a solution to this problem: The idea of inborn concepts is inconsistent with modern biology and glosses over the process of their development, which is the job of a developmental account to explain. Even though all three frameworks have been argued to account for empirical data, they remain untenable in light of the above theoretical issues. Although needing much work, the alternative, action-based perspective we have presented offers a framework that naturally avoids foundationalism and nativism.

Interpretations of the empirical data (infant mindreading, context-sensitive developmental progressions, and cross-cultural differences in folk psychology) differ greatly between the two paradigms. Foundational concepts set the course for development in the traditional frameworks, which means that experience plays only a mediating role in that ontologically internal composition. The action-based paradigm, on the other hand, adopts a radically different, grounded position, and sees past experience as *constitutive* of representations involved in social cognition. Variance in social cognitive skills observed across cultures, and other linguistic and social contexts, is therefore deeply significant as it evidences *genuine* conceptual differences in people from different socio-linguistic contexts. Different interactions that those contexts afford lead to the emergence of essentially different representations of the human social world.

Acknowledgements We would like to thank Jedediah WP Allen, Him Cheung, Marcin Miłkowski, Delphine Wartelle, Julia Wolf, and three anonymous reviewers for their useful comments and discussion on earlier versions of this article.

Funding This paper is supported by the Polish National Science Centre's Grant: HARMONIA 6 (UMO-2014/14/M/HS1/00436) for the research project entitled "The involvement of language, culture-specific factors and folk intuitions in mindreading activity and social cognition".

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Allen, J. W. P., & Bickhard, M. H. (2013). Stepping off the pendulum: Why only an action-based approach can transcend the nativist–empiricist debate. *Cognitive Development, 28*(2), 96–133. https://doi.org/10.1016/j.cogdev.2013.01.002.

Andrews, K. (2012). *Do apes read minds?: Toward a new folk psychology/Kristin Andrews*. London: MIT Press.

Apperly, I. A. (2012a). *Mindreaders: The cognitive basis of 'theory of mind'*. Hove: Psychology.

Apperly, I. A. (2012b). What is "theory of mind"? Concepts, cognitive processes and individual differences. *Quarterly Journal of Experimental Psychology, 65*(5), 825–839. https://doi.org/10.1080/17470218.2012.676055.

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116*(4), 953–970. https://doi.org/10.1037/a0016923.

Astington, J. W., & Baird, J. A. (Eds.). (2005). *Why language matters for theory of mind*. New York, NY: Oxford University Press.

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology, 35*(5), 1311–1320. https://doi.org/10.1037//0012-1649.35.5.1311.

Banovsky, J. (2016). Theories, structures and simulations in the research of early mentalizing. *Cognitive Systems Research, 40,* 129–143. https://doi.org/10.1016/j.cogsys.2016.05.003.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59,* 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639.

Bickhard, M. H. (1998). Levels of representationality. *Journal of Experimental and Theoretical Artificial Intelligence, 10*(2), 179–215. https://doi.org/10.1080/095281398146824.

Bickhard, M. H. (2007). Language as an interaction system. *New Ideas in Psychology, 25*(2), 171–187. https://doi.org/10.1016/j.newideapsych.2007.02.006.

Bickhard, M. H. (2009a). Interactivism. In J. Symons & P. Calvo (Eds.), *The Routledge companion to philosophy of psychology* (pp. 345–359). London, NY: Routledge Taylor and Francis Group.

Bickhard, M. H. (2009b). The interactivist model. *Synthese, 166*(3), 547–591. https://doi.org/10.1007/s11229-008-9375-x.

Bickhard, M. H. (2010). Interactive knowing: The metaphysics of intentionality. In R. Poli & J. Seibt (Eds.), *Theory and applications of ontology: Philosophical perspectives* (pp. 207–229). London: Springer.

Bickhard, M. H., & Terveen, L. (1995). *Foundational issues in artificial intelligence and cognitive science: Impasse and solution. Advances in psychology* (Vol. 109). Amsterdam, NY: Elsevier.

Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind and Language, 28*(5), 606–637. https://doi.org/10.1111/mila.12036.

Cain, M. J. (2013). *Fodor: Language, mind and philosophy*. *Key contemporary thinkers*. Hoboken: Wiley. Retrieved from http://gbv.eblib.com/patron/FullRecord.aspx?p=1215760. Accessed Dec 15, 2017.

Campbell, R. L., & Bickhard, M. H. (1986). *Knowing levels and developmental stages. Contributions to human development* (Vol. 16). Basel: S. Karger.

Carpendale, J. I. M., Hammond, S. I., & Atwood, S. (2013). A relational developmental systems approach to moral development. In R. M. Lerner & J. B. Benson (Eds.), *Advances in child development and behavior. Embodiment and epigenesis: Theoretical and methodological issues in understanding the role of biology within the relational developmental system—Part B: Ontogenetic dimensions* (Vol. 45, pp. 125–153). Amsterdam: Elsevier. https://doi.org/10.1016/B978-0-12-397946-9.00006-3.

Carpendale, J. I. M., & Lewis, C. (2004). Constructing an understanding of mind: The development of children's social understanding within social interaction. *The Behavioral and Brain Sciences, 27*(1), 79–96. **discussion 96-151**.

Carpendale, J. I. M., & Lewis, C. (2006). *How children develop social understanding. Understanding children's worlds*. Malden, MA: Blackwell Publishers.

Carpendale, J. I. M., & Wereha, T. J. (2013). Understanding common developmental timetables across cultures from a developmental systems perspective. *Human Development, 56*(3), 207–212. https://doi.org/10.1159/000351276.

Carruthers, P. (2013). Mindreading in Infancy. *Mind and Language, 28*(2), 141–172. https://doi.org/10.1111/mila.12014.

Carruthers, P. (2015). Mindreading in adults: Evaluating two-systems views. *Synthese*. https://doi.org/10.1007/s11229-015-0792-3.

Carruthers, P. (2016). Two systems for mindreading? *Review of Philosophy and Psychology, 7*(1), 141–162. https://doi.org/10.1007/s13164-015-0259-y.

Cheung, H., Hsuan-Chih, C., Creed, N., Ng, L., Ping, W. S., & Mo, L. (2004). Relative roles of general and complementation language in theory-of-mind development: Evidence from Cantonese and English. *Child Development, 75*(4), 1155–1170. https://doi.org/10.1111/j.1467-8624.2004.00731.x.

Chapman, M. (1999). Constructivism and the problem of reality. *Journal of Applied Developmental Psychology, 20*, 31–43. https://doi.org/10.1016/S0193-3973(99)80002-1
Churchland, P. M. (1998). Folk psychology. In P. M. Churchland & P. S. Churchland (Eds.), *Bradford Bks. On the contrary: Critical essays, 1987–1997* (pp. 3–15). Cambridge, Mass: MIT Press.
Clark, A. (1997). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.
Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford, NY: Oxford University Press.
Davies, M., & Stone, T. (2001). Mental simulation, tacit theory, and the threat of collapse. *Philosophical Topics, 29*(1), 127–173. https://doi.org/10.5840/philtopics2001291/212.
De Bruin, L. C., & Newen, A. (2012). An association account of false belief understanding. *Cognition, 123*(2), 240–259. https://doi.org/10.1016/j.cognition.2011.12.016.
De Villiers, P. A. (2005). The role of language in theory of mind development: What deaf children tell us. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 266–297). New York, Oxford: Oxford University Press.
De Villiers, J. G., & de Villiers, P. A. (2009). Complements enable representation of the contents of false beliefs: The evolution of a theory of theory of mind. In S. H. Foster-Cohen (Ed.), *Palgrave advances in linguistics. Language acquisition* (pp. 169–195). Houndmills: Palgrave Macmillan. https://doi.org/10.1057/9780230240780_8.
De Villiers, J. G., & de Villiers, P. A. (2014). The role of language in theory of mind development. *Topics in Language Disorders, 34*(4), 313–328. https://doi.org/10.1097/TLD.0000000000000037.
De Villiers, J. G., & Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief-understanding. *Cognitive Development, 17*(1), 1037–1060. https://doi.org/10.1016/S0885-2014(02)00073-4.
Dennett, D. C. (1996). *Kinds of minds: Toward an understanding of consciousness. Science masters*. New York, NY: BasicBooks.
Devine, R. T., & Hughes, C. (2016). Family correlates of false belief understanding in early childhood: A meta-analysis. *Child Development*. https://doi.org/10.1111/cdev.12682.
Dixson, H. G. W., Komugabe-Dixson, A. F., Dixson, B. J., & Low, J. (2017). Scaling theory of mind in a small-scale society: A case study from Vanuatu. *Child Development*. https://doi.org/10.1111/cdev.12919.
Elman, J. L. (Ed.). (1996). *Neural network modeling and connectionism. Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.
Engel, A. K., Friston, K. J., & Kragic, D. (Eds.). (2015). *Strüngmann forum reports. The pragmatic turn: Toward action-oriented views in cognitive science*. Cambridge, MA: The MIT Press.
Fenici, M. (2012). Embodied social cognition and embedded theory of mind. *Biolinguistics, 6*(3–4), 276–307.
Fenici, M. (2013). Social cognitive abilities in infancy: Is mindreading the best explanation? *Philosophical Psychology, 28*(3), 387–411. https://doi.org/10.1080/09515089.2013.865096.
Fenici, M. (2015). A simple explanation of apparent early mindreading: Infants' sensitivity to goals and gaze direction. *Phenomenology and the Cognitive Sciences, 14*(3), 497–515. https://doi.org/10.1007/s11097-014-9345-3.
Fenici, M. (2017). What is the role of experience in children's success in the false belief test: Maturation, facilitation, attunement or induction? *Mind and Language, 32*(3), 308–337. https://doi.org/10.1111/mila.12145.
Fenici, M., & Garofoli, D. (2017). The biocultural emergence of mindreading: Integrating cognitive archaeology and human development. *Journal of Cultural Cognitive Science, 1*(2), 89–117. https://doi.org/10.1007/s41809-017-0008-0.
Fodor, J. A. (1975). *The language of thought. The language and thought series*. New York, NY: Crowell.
Fodor, J. A. (1983). *The modularity of mind*. Cambridge: The MIT Press. Retrieved from http://gbv.eblib.com/patron/FullRecord.aspx?p=3339779. Accessed Dec 15, 2017.
Fodor, J. A. (1992). A theory of the child's theory of mind. *Cognition, 44*(3), 283–296. https://doi.org/10.1016/0010-0277(92)90004-2.
Furrow, D., Moore, C., Davidge, J., & Chiasson, L. (1992). Mental terms in mothers' and children's speech: Similarities and relationships. *Journal of Child Language, 19*(03), 617. https://doi.org/10.1017/S0305000900011594.
Galende, N., Sánchez de Miguel, M., & Arranz Freijo, E. B. (2014). Family context and theory of mind development in early childhood. In O. N. Saracho (Ed.), *Contemporary perspectives in early childhood*

education. Contemporary perspectives on research in theory of mind in early childhood education (pp. 365–380). Charlotte, NC: Information Age Publishing.

Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition, 17*(2), 535–543. https://doi.org/10.1016/j.concog.2008.03.003.

Gallagher, S., & Hutto, D. D. (2008). Understanding others through primary interaction and narrative practice. In J. Zlatev, T. P. Racine, C. Sinha, & E. Itkonen (Eds.), *Converging evidence in language and communication research. The shared mind* (Vol. 12, pp. 17–38). Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/celcr.12.04gal.

Goodman, N. D., Baker, C. L., Bonawitz, E. B., Mansinghka, V. K., Gopnik, A., Wellman, H. M., et al. (2006). Intuitive theories of mind: A rational approach to false belief. In *Proceedings of the CogSci 2006, 28th annual conference of the cognitive science society in cooperation with the 5th international conference of the cognitive science: July 26–29, 2006, Vancouver, BC, Canada*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gopnik, A., Meltzoff, A. N., & Kuhl, P. K. (1999). *The scientist in the crib: What early learning tells us about the mind*. Pymble, NSW: HarperCollins e-books.

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind and Language, 7*(1–2), 145–171. https://doi.org/10.1111/j.1468-0017.1992.tb00202.x.

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

Gould, S. J. (1991). Exaptation: A crucial tool for an evolutionary psychology. *Journal of Social Issues, 47*(3), 43–65. https://doi.org/10.1111/j.1540-4560.1991.tb01822.x.

Gould, S. J., & Lewontin, R. C. (1979). The Spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society B: Biological Sciences, 205*(1161), 581–598. https://doi.org/10.1098/rspb.1979.0086.

Gould, S. J., & Vrba, E. S. (1982). Exaptation—A missing term in the science of form. *Paleobiology, 8*(01), 4–15. https://doi.org/10.1017/S0094837300004310.

Gut, A., & Mirski, R. (2016). In search of a theory: The interpretative challenge of empirical findings on cultural variance in mindreading. *Studies in Logic and Grammar and Rhetoric, 48*(61), 201–230. https://doi.org/10.1515/slgr-2016-0063.

Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science, 6*(3), 346–359. https://doi.org/10.1111/1467-7687.00289.

Helming, K. A., Strickland, B., & Jacob, P. (2014). Making sense of early false-belief understanding. *Trends in Cognitive Sciences, 18*(4), 167–170. https://doi.org/10.1016/j.tics.2014.01.005.

Helming, K. A., Strickland, B., & Jacob, P. (2016). Solving the puzzle about early belief-ascription. *Mind and Language, 31*(4), 438–469. https://doi.org/10.1111/mila.12114.

Heyes, C. (2014). False belief in infancy: A fresh look. *Developmental Science, 17*(5), 647–659. https://doi.org/10.1111/desc.12148.

Howell, S. (1981). Rules Not Words. In P. Heelas & A. Lock (Eds.), *Language, thought, and culture. Indigenous psychologies: The anthropology of the self* (pp. 133–143). London: Academic Press.

Hutto, D. D. (2008). *Folk psychological narratives: The sociocultural basis of understanding reasons*. Cambridge, MA: MIT Press.

Hutto, D. D. (2015). Basic social cognition without mindreading: Minding minds without attributing contents. *Synthese, 194*(3), 827–846. https://doi.org/10.1007/s11229-015-0831-0.

Hutto, D. D., & Myin, E. (2013). *Radicalizing enactivism: Basic minds without content*. Cambridge, MA: MIT Press.

Hutto, D. D., & Myin, E. (2017). *Evolving enactivism: Basic minds meet content*. Cambridge, MA: MIT Press.

Jenkins, J. M., & Astington, J. W. (1996). Cognitive factors and family structure associated with theory of mind development in young children. *Developmental Psychology, 32*(1), 70–78. https://doi.org/10.1037/0012-1649.32.1.70.

Kallberg-Schroff, M., & Miller, J. M. (2014). Culture and the development of Theory of Mind. In O. N. Saracho (Ed.), *Contemporary perspectives in early childhood education. Contemporary perspectives on research in theory of mind in early childhood education* (pp. 343–364). Charlotte, NA: Information Age Publishing.

Kobayashi, C., Glover, G. H., & Temple, E. (2006). Cultural and linguistic influence on neural bases of 'Theory of Mind': An fMRI study with Japanese bilinguals. *Brain and Language, 98*(2), 210–220. https://doi.org/10.1016/j.bandl.2006.04.013.

Kristen, S., & Sodian, B. (2014). Theory of mind (ToM) in early education: Developmental progression of early theory of mind skills, social developmental factors and the importance of ToM for learning. In O. N. Saracho (Ed.), *Contemporary perspectives in early childhood education. Contemporary perspectives on research in theory of mind in early childhood education* (pp. 291–320). Charlotte, NC: Information Age Publishing.

Kuntoro, I. A., Peterson, C. C., & Slaughter, V. P. (2017). Culture, parenting, and children's theory of mind development in Indonesia. *Journal of Cross-Cultural Psychology, 48*(9), 1389–1409. https://doi.org/10.1177/0022022117725404.

Lebra, T. S. (1993). Culture, self, and communication in Japan and the United States. In W. B. Gudykunst (Ed.), *SUNY series, human communication processes. Communication in Japan and the United States* (pp. 51–87). Albany, NY: State University of New York.

Leslie, A. M. (1994). ToMM, ToBy, and agency: Core architecture and domain specificity. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture* (pp. 119–148). Cambridge: Cambridge University Press.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in "theory of mind". *Trends in Cognitive Sciences, 8*(12), 528–533. https://doi.org/10.1016/j.tics.2004.10.001.

Lewis, C., Carpendale, J. I. M., & Stack, J. (2013). Anticipation and social interaction: Commentary on "Stepping off the pendulum: Why only an action-based approach can transcend the nativist–empiricist debate" by J. Allen and M. Bickhard. *Cognitive Development, 28*(2), 159–163. https://doi.org/10.1016/j.cogdev.2013.01.008.

Lewkowicz, D. J. (2011). The biological implausibility of the nature–nurture dichotomy and what it means for the study of infancy. *Infancy, 16*(4), 331–367. https://doi.org/10.1111/j.1532-7078.2011.00079.x.

Lewontin, R. C. (2001). *The triple helix: Gene, organism and environment*. Cambridge, MA: Harvard University Press.

Lightfoot, D. (1989). The child's trigger experience: Degree-0 learnability. *Behavioral and Brain Sciences, 12*(02), 321. https://doi.org/10.1017/S0140525X00048883.

Lillard, A. S. (1998). Ethnopsychologies: Cultural variations in theories of mind. *Psychological Bulletin, 123*(1), 3–32. https://doi.org/10.1037/0033-2909.123.1.3.

Low, J., Apperly, I. A., Butterfill, S. A., & Rakoczy, H. (2016). Cognitive architecture of belief reasoning in children and adults: A primer on the two-systems account. *Child Development Perspectives, 10*(3), 184–189. https://doi.org/10.1111/cdep.12183.

Mameli, M., & Bateson, P. (2011). An evaluation of the concept of innateness. *Philosophical Transactions of the Royal Society of London: Series B, Biological Sciences, 366*(1563), 436–443. https://doi.org/10.1098/rstb.2010.0174.

Marr, D. (1982/2010). *Vision: A computational investigation into the human representation and processing of visual information*. Cambridge, MA: MIT Press.

Mayer, A., & Träuble, B. (2012). Synchrony in the onset of mental state understanding across cultures?: A study among children in Samoa. *International Journal of Behavioral Development, 37*(1), 21–28. https://doi.org/10.1177/0165025412454030.

Meristo, M., Morgan, G., Geraci, A., Iozzi, L., Hjelmquist, E., Surian, L., et al. (2012). Belief attribution in deaf and hearing infants. *Developmental Science, 15*(5), 633–640. https://doi.org/10.1111/j.1467-7687.2012.01155.x.

Meristo, M., Strid, K., & Hjelmquist, E. (2016). Early conversational environment enables spontaneous belief attribution in deaf children. *Cognition, 157,* 139–145. https://doi.org/10.1016/j.cognition.2016.08.023.

Milligan, K., Astington, J. W., & Dack, La. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development, 78*(2), 622–646. https://doi.org/10.1111/j.1467-8624.2007.01018.x.

Mills, S. (2001). The idea of different folk psychologies. *International Journal of Philosophical Studies, 9*(4), 501–519. https://doi.org/10.1080/09672550110081285.

Naito, M. (2014). From theory of mind to theory of relation: Sociocultural perspectives on Japanese children's social understanding. In O. N. Saracho (Ed.), *Contemporary perspectives in early childhood education. Contemporary perspectives on research in theory of mind in early childhood education* (pp. 381–408). Charlotte, NA: Information Age Publishing.

Nawaz, S., Hanif, R., & Lewis, C. (2014). 'Theory of mind' development of Pakistani children: Do preschoolers acquire an understanding of desire, pretence and belief in a universal sequence? *European Journal of Developmental Psychology, 12*(2), 177–188. https://doi.org/10.1080/17405629.2014.973843.

Nelson, K. (2005). Language pathways into the community of minds. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 26–49). New York, NY: Oxford University Press.

Nelson, K. (2007). *Young minds in social worlds: Experience, meaning, and memory*. Cambridge, MA: Harvard University Press.

Newcombe, N. S. (2011). What is neoconstructivism? *Child Development Perspectives, 5*(3), 157–160. https://doi.org/10.1111/j.1750-8606.2011.00180.x.

Newen, A. (2015). Understanding others: The person model theory. In T. Metzinger & J. M. Windt (Eds.), *Open MIND: 26(T)*. Frankfurt am Main: MIND Group. https://doi.org/10.15502/9783958570320.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences, 24*(05), 939–973. https://doi.org/10.1017/S0140525X01000115.

Olson, D. R. (1989). On the origins of beliefs and other intentional states in children. In J. W. Astington (Ed.), *Developing theories of mind: [Papers from the International Conference on Developing Theories of Mind, University of Toronto, May 1986 …]* (1st ed., pp. 414–426). Cambridge: Cambridge University Press.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science (New York, NY), 308*(5719), 255–258. https://doi.org/10.1126/science.1107621.

Oyama, S. (1985/2000). *The ontogeny of information*. Durham: Duke University Press.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology, 5*(2), 125–137. https://doi.org/10.1111/j.2044-835X.1987.tb01048.x.

Perner, J., & Ruffman, T. (2005). Psychology. Infants' insight into the mind: How deep? *Science (New York, NY), 308*(5719), 214–216. https://doi.org/10.1126/science.1111656.

Perner, J., Ruffman, T., & Leekam, S. R. (1994). Theory of mind is contagious: You catch it from your sibs. *Child Development, 65*(4), 1228–1238. https://doi.org/10.1111/j.1467-8624.1994.tb00814.x.

Peskin, J., & Astington, J. W. (2004). The effects of adding metacognitive language to story texts. *Cognitive Development, 19*(2), 253–273. https://doi.org/10.1016/j.cogdev.2004.01.003.

Pezzulo, G., Vosgerau, G., Frith, U., Hamilton, A. F., Heyes, d. C., & Iriki, A. (2015). Acting up: An approach to the study of cognitive development. In A. K. Engel, K. J. Friston, & D. Kragic (Eds.), *Strüngmann forum reports. The pragmatic turn: Toward action-oriented views in cognitive science* (pp. 49–77). Cambridge, MA: The MIT Press.

Pigliucci, M., & Müller, G. B. (Eds.). (2010). *Evolution—The extended synthesis*. Cambridge, MA: The MIT Press.

Pinker, S. (2014/1994). *The language instinct: How the mind creates language*. Bloomberg: Brilliance Audio.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(04), 515. https://doi.org/10.1017/S0140525X00076512.

Pyers, J. E., & Senghas, A. (2009). Language promotes false-belief understanding: Evidence from learners of a new sign language. *Psychological Science, 20*(7), 805–812. https://doi.org/10.1111/j.1467-9280.2009.02377.x.

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.

Quine, W. V. (1960/2013). *Word and object* (New edition). Cambridge, MA: MIT Press.

Racine, T. P. (2013). How useful are the concepts "innate" and "adaptation" for explaining human development. *Human Development, 56*(3), 141–146. https://doi.org/10.1159/000350925.

Ruffman, T. (2014). To belief or not belief: Children's theory of mind. *Developmental Review, 34*(3), 265–293. https://doi.org/10.1016/j.dr.2014.04.001.

Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children's and mothers' mental state language and theory-of-mind understanding. *Child Development, 73*(3), 734–751.

Ruffman, T., Slade, L., Rowlandson, K., Rumsey, C., & Garnham, A. (2003). How language relates to belief, desire, and emotion understanding. *Cognitive Development, 18*(2), 139–158. https://doi.org/10.1016/S0885-2014(03)00002-9.

Ruffman, T., & Taumoepeau, M. (2014). When and how does a theory of mind arise? In O. N. Saracho (Ed.), *Contemporary perspectives in early childhood education. Contemporary perspectives on research in theory of mind in early childhood education* (pp. 45–63). Charlotte, NC: Information Age Publishing.

Samuels, R. (2002). Nativism in Cognitive Science. *Mind & Language, 17*, 233–265. https://doi.org/10.1111/1468-0017.00197

Scholl, B. J., & Leslie, A. M. (1999). Modularity, development and 'theory of mind'. *Mind and Language, 14*(1), 131–153. https://doi.org/10.1111/1468-0017.00106.

Shahaeian, A., Nielsen, M., Peterson, C. C., & Slaughter, V. (2014). Iranian mothers' disciplinary strategies and theory of mind in children. *Journal of Cross-Cultural Psychology, 45*(7), 1110–1123. https://doi.org/10.1177/0022022114534772.

Shahaeian, A., Peterson, C. C., Slaughter, V. P., & Wellman, H. M. (2011). Culture and the sequence of steps in theory of mind development. *Developmental Psychology, 47*(5), 1239–1247. https://doi.org/10.1037/a0023899.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts. Cognitive science series* (Vol. 4). Cambridge, MA: Harvard University Press.

Spencer, J. P., Blumberg, M. S., McMurray, B., Robinson, S. R., Samuelson, L. K., & Tomblin, J. B. (2009). Short arms and talking eggs: Why we should no longer abide the nativist–empiricist debate. *Child Development Perspectives, 3*(2), 79–87. https://doi.org/10.1111/j.1750-8606.2009.00081.x.

Strijbos, D. W., & De Bruin, L. C. (2013). Universal belief-desire psychology?: A dilemma for theory and simulation theory. *Philosophical Psychology, 26*(5), 744–764. https://doi.org/10.1080/09515089.2012.711034.

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18*(7), 580–586. https://doi.org/10.1111/j.1467-9280.2007.01943.x.

Tafreshi, D., & Racine, T. P. (2016). Children's interpretive theory of mind: The role of mothers' personal epistemologies and mother–child talk about interpretation. *Cognitive Development, 39,* 57–70. https://doi.org/10.1016/j.cogdev.2016.04.003.

Tager-Flusberg, H., & Joseph, R. M. (2005). How language facilitates the acquisition of false-belief understanding in children with autism. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 298–318). New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195159912.003.0014.

Thelen, E., & Smith, L. B. (2002/1996). *A dynamic systems approach to the development of cognition and action (5. print). A Bradford book*. Cambridge, MA: MIT Press.

Träuble, B., Marinović, V., & Pauen, S. (2010). Early theory of mind competencies: Do infants understand others' beliefs? *Infancy, 15*(4), 434–444. https://doi.org/10.1111/j.1532-7078.2009.00025.x.

Van Gelder, T., & Smylie, J. (1995). What might cognition be, if not computation? *Journal of Philosophy, 92*(7), 345–381. https://doi.org/10.2307/2941061.

Venter, C., & Cohen, D. (2004). The century of biology. *New Perspectives Quarterly, 21*(4), 73–77. https://doi.org/10.1111/j.1540-5842.2004.00701.x.

Vinden, P. G. (1996). Junin Quechua children's understanding of mind. *Child Development, 67*(4), 1707–1716. https://doi.org/10.1111/j.1467-8624.1996.tb01822.x.

Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy, 11*(3), 271–294. https://doi.org/10.1111/j.1532-7078.2007.tb00227.x.

Wellman, H. M. (2014). *Making minds: How theory of mind develops. Oxford series in cognitive development*. Oxford, NY: Oxford University Press.

Wellman, H. M., Fang, F., & Peterson, C. C. (2011). Sequential progressions in a theory-of-mind scale: Longitudinal perspectives. *Child Development, 82*(3), 780–792. https://doi.org/10.1111/j.1467-8624.2011.01583.x.

Wellman, H. M., & Peterson, C. C. (2013). Theory of mind, development, and deafness. In S. Baron-Cohen, H. Tager-Flusberg, & M. Lombardo (Eds.), *Understanding other minds: Perspectives from developmental social neuroscience* (3rd ed., pp. 51–71). Oxford: Oxford University Press.

Westra, E., & Carruthers, P. (2017). Pragmatic development explains the theory-of-mind scale. *Cognition, 158,* 165–176. https://doi.org/10.1016/j.cognition.2016.10.021.

Wierzbicka, A. (1992). *Semantics, culture, and cognition: Universal human concepts in culture-specific configurations*. New York, NY: Oxford University Press.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5.

Woolfe, T., Want, S. C., & Siegal, M. (2002). Signposts to development: Theory of mind in deaf children. *Child Development, 73*(3), 768–778. https://doi.org/10.1111/1467-8624.00437.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature, 358*(6389), 749–750. https://doi.org/10.1038/358749a0.

Zawidzki, T. (2013). *Mindshaping: A new framework for understanding human social cognition*. Cambridge, MA: MIT Press.

New Ideas in Psychology 62 (2021) 100856

Contents lists available at ScienceDirect

New Ideas in Psychology

journal homepage: http://www.elsevier.com/locate/newideapsych

Conventional minds: An interactivist perspective on social cognition and its enculturation

Robert Mirski [a,b,*], Mark H. Bickhard [c]

[a] Department of Philosophy, The John Paul II Catholic University of Lublin Al, Racławickie 14, 20-950, Lublin, Poland
[b] Nicolaus Copernicus University in Toruń, ul. Gagarina 11, 87-10,0, Torun, Poland
[c] Department of Philosophy, Lehigh University, 27 Memorial Drive West, Bethlehem, PA, 18015, USA

ARTICLE INFO

Keywords:
Social cognition
Theory of mind
Mindreading
Folk psychology
Enculturation
Interactivism

ABSTRACT

We argue that the traditional theory of mind models of social cognition face in-principle problems in accounting for enculturation of social cognition, and offer an alternative model advanced within the interactivist framework. In the critical section, we argue that theory of mind accounts' encodingist model of mental representation renders them unable to account for enculturation. We focus on the three problems: (1) the copy problem and impossibility of internalization; (2) foundationalism and the impossibility of acquisition of culturally specific content; and (3) the frame problems and the inadequacy of mental-state attribution as a way of coordinating social interaction among (encultured) individuals. The positive section begins with a brief sketch of the theoretical basics of interactivism, followed by a more focused presentation of the interactivist model of social cognition, and concludes with a discussion of a number of issues most widely debated in the social cognition literature.

1. Introduction

Philosophy and social sciences have been interested in the subject of social cognition since their very conception. Over recent years, however, the problem has been largely dominated by the area of research called theory of mind (ToM). Originating a couple decades back (Premack & Woodruff, 1978), the idea that human sociality is founded on the ability to attribute mental states to others has dominated psychology and related fields (reviews of the field can be found in Baron-Cohen, Tager-Flusberg, & Lombardo, 2013; Fenici, 2017a; Wellman, 2018). So much so that thoughts on the subject coming from other research traditions have to be laboriously reintroduced into this ToM-dominated climate today (e.g. Carpendale & Lewis, 2006; De Jaegher & Di Paolo, 2007; Gallagher, 2001, 2008; Nelson, 2007). The problem of culture's role in how we view one another fits this trend: While the relation between culture and the individual is one of the perennial topics in social sciences at large, it is only over the recent years that theory of mind research has begun to explore culture's significance for socio-cognitive abilities and their development (Lavelle, 2019; Taumoepeau, 2019).

Most researchers working within ToM do not consider culture exceptionally problematic for their models. Although data are accumulating of a cross-cultural variance in performance on standard ToM tasks both among children (Dixson, Komugabe-Dixson, Dixson, & Low, 2017; Hughes et al., 2014; Mayer & Träuble, 2012; Vinden, 1999, 2002; Wellman, Fang, Liu, Zhu, & Liu, 2006; Wellman & Liu, 2004) as well as adults (Adams et al., 2010; Kobayashi, Glover, & Temple, 2006; Mitchell, Macrae, & Banaji, 2006; Perez-Zapata, Slaughter, & Henry, 2016; Savitsky, Keysar, Epley, Carter, & Swanson, 2011), ToM theoreticians tend to assume that they are simply evidence of theory of mind – a mental attribution mechanism – developing differentially depending on the cultural context. Alternatively, those of more nativist leanings argue that culture modulates other cognitive process while the ToM mechanism is innate and impervious to the impact of culture (Helming, Strickland, & Jacob, 2014, 2016; Westra & Carruthers, 2017).

In this paper, we point out that ToM accounts of cultural effects on social cognition are far from being free of controversy. We argue that ToM models cannot in fact account for genuine enculturation of social cognition – much less for the existence of a social reality – due to fatal problems intrinsic to their overarching theoretical framework. As an alternative, we present an account of social cognition within the interactivist framework (Bickhard, 2009) and demonstrate how it accommodates enculturation of social cognition and emergence of social

* Corresponding author. The John Paul II Catholic University of Lublin Al. Racławickie 14, 20-950, Lublin, Poland.
E-mail address: robmirski@gmail.com (R. Mirski).
URL: https://www.researchgate.net/profile/Robert_Mirski (R. Mirski).

https://doi.org/10.1016/j.newideapsych.2021.100856
Received 26 June 2020; Received in revised form 7 February 2021; Accepted 8 February 2021

reality. Finally, we discuss the proposed framework in the context of some of the most discussed problems in the ToM-dominated social cognition research: false belief understanding, folk psychology, cross-cultural differences in social cognition, and the influence of language on socio-cognitive development.

2. The uncultured ToM

ToM models come in a few variants, the most notable of which in the present context are nativism (Carruthers, 2013, 2015; Fodor, 1992; Leslie, Friedman, & German, 2004; Scott & Baillargeon, 2017; Westra, 2017), and rational constructivism (theory theory) (Gopnik, 2011; Gopnik & Wellman, 1992; Meltzoff & Gopnik, 2013; Wellman, 2014).[1] The underlying presupposition that unites these otherwise divergent views is that social cognition is underwritten by a mental-state attribution mechanism – the titular ToM. The mechanism is claimed to be indispensable in competent social interaction; it is thanks to sub-personal mindreading (i.e. not folk psychological, but of lower-level of organization) – as the workings of ToM are often termed – that we can imitate, learn a language, feel empathy, and essentially do any other thing that involves participation in social interaction.

This idea has been criticized extensively from an empirical perspective; many researchers have pointed out that data currently available speak against the ToM-mechanism views (e.g. Carpendale & Lewis, 2006; Fenici & Garofoli, 2017; Heyes, 2018; Nelson, 2007). As this paper focuses on theoretical issues, we will not review these debates. We do, however, believe that our positive proposal accounts for the data problematic to ToM models, which we will point out where relevant. In this section, we offer criticism of the ToM accounts' underlying theoretical framework and argue that it faces in-principle problems when the issue of enculturation of social cognition is concerned.

All traditional ToM proposals are characterized by encodingism (see, e.g. Bickhard & Terveen, 1995; Brette, 2019; Mirski & Bickhard, 2019). Encodingism is the view that the fundamental form of representation is via encoded meanings. That is, according to encodingism, cognition consists in minds utilizing mental representations that refer to the world via their encoded content, which predicates the world to be a particular way (e.g. if a representation of a car is tokened, the organism predicates that there is a car in the field of vision). Encoding relationships can be rendered in a number of ways – as correspondence, correlation, informational relationship, covariance etc. – but they all boil down to the same principle of the state of the representational vehicle "standing in" for what is being represented. This harkens back to the classical, computational model of the mind (e.g. Fodor, 1975; Newell & Simon, 2006).[2] Perception is transducing or encoding incoming stimuli into appropriate symbols, and action is transducing computed representations back into motor directions. It is now the standard framework of cognitive psychology (e.g. Groome & Brace, 2014) and a taken for granted truth in ToM proposals.

Encodingism births a plethora of problems (Bickhard, 2009; Bickhard & Terveen, 1995; Brette, 2019; Mirski & Bickhard, 2019), but the ones most relevant for the present discussion are the copy problem, foundationalism, and the frame problems, which we discuss below.

2.1. The copy problem and internalization

A fundamental characteristic of encodings is that by their very nature, they require an interpreter to have their meaning; an encoded representation has to inform someone or something's functioning in a way that is consistent with what the representation encodes, what it stands for. It is the interpreter that provides content for the representational vehicle.

One fatal consequence of the necessity of an interpreter is that the organism cannot, in fact, encode any new fundamental meanings. The only access to reality that the organism has is through encodings and so any 'new' knowledge would have to already be known (to the interpreter). An interpreter must interpret into something already available. Consequently, if we model our cognitive system as a system that uses solely encoded representations, then we make it impossible for it to learn new representations – to encode new meanings.

The above problem has been pointed out already by Piaget (1971) in his copy argument. If our representations of the world are copies of it in some sense (encodings), then we would have to already know the thing that we wish to copy or encode. The sensory "imprint" of an object (e.g. chair) carries no information about what it is an imprint of, and so to take it as a representation of a chair requires the organism to already know what chairs are. In other words, encoding new meanings would require some independent epistemic access to that which is to be copied, which is impossible if encodings are taken to be the fundamental type of representation.

Relevant to our present purposes, the copy problem has serious consequences for the notion of internalization as a model of enculturation. Both Piaget (1999[1951]; Piaget & Inhelder, 2000[1969]) and Vygotsky (1978) used the notion of internalization, but in the context of enculturation, it is Vygotsky's proposal that is most relevant and has had the most influence (Zittoun & Gillespie, 2015). While there have been various elaborations of the model, the fundamental problem is that it simply cannot happen as long as we take cognition to be based on encoded representations. The principal idea behind internalization is that something that is external – e.g. a social rule or norm or culturally specific mental concept – is transmitted or impressed into the child's mind (copied). Within an encodingist framework, however, the only way any internalization could possibly happen would be by encoding new content. But this is precisely what we have shown is impossible – the child would have to already know the rule or norm or concept in order to internalize it (for more exhaustive criticism, see Christopher & Bickhard, 2007).[3]

2.2. Foundationalism and enculturation

There is (seemingly) one way out of the copy problem. If the system came with a set of already existing representations with encoded meanings known by the system innately, then new representations could be created as recombinations of these innate foundations – no "copying" would be necessary. In fact, an innate representational foundation is the only way out for encodingist models to account for learning, and we refer to the theoretical necessity of postulating such foundations as *foundationalism* (Allen & Bickhard, 2013; Mirski & Gut, 2018). The concept of the chair cannot be "copied" from experience, but it can be assembled on its basis if the organism has already available its

[1] There have naturally been others, such as simulation theory (Goldman, 2008) or Perner's teleology and mental files (e.g. Doherty & Perner, 2020; Perner, Priewasser, & Roessler, 2018), but their role in debates on culture's role in socio-cognitive development has been much lesser than the two mentioned. In any case, as our criticism is against the underlying theoretical framework of ToM models, any other accounts are its subject just the same insofar as they follow the same presupposition that social cognition is founded on the ability of mental-state attribution (albeit differing in detail, both simulationism and Perner's proposals do take that view as well).

[2] And in some version even to the ancient Greek idea that "like represents like" – see Aristotle's signet ring argument.

[3] Nelson (2007, pp. 68–71) argues that the copy (encodingist) reading of Vygotsky's internalization is incorrect, and that an action-based constructivist interpretation of it, much like the proposal offered in this text, is a more accurate exegesis of his ideas. This might be true, but the fact remains that internalization within encodingism is untenable, and even if not originally intended as such, the term invites such a reading.

constituent representations – experience is then viewed as instructions for how to combine whatever innate representations the particular model starts with. The difference between encodingist rationalists (or concept nativists) and empiricists is in essence that the former start with fairly complex foundational representations, while the latter posit more primitive building blocks. In both cases, however, development necessarily boils down to reconfiguration of the representational givens. It is customary for encodingist models to be open about their foundations, as we see explicitly done by both nativism, as well as theory theory (Gopnik, 2003; Wellman, 2014, p. 197). Even when it is not openly claimed, however, foundationalism is a necessary constraint on all encodingist models of cognition – these accounts have to rely on an idea of cognitive development as reconfiguration of the innate representations since new ones cannot *in principle* be created.[4]

Central to our present purpose, foundationalism does not really help with the problem of enculturation. If no genuine learning is possible, then no genuine cultural learning is possible. Within an encodingist framework, children start with the foundational concepts, form hypotheses, some of which are confirmed in some cultures, but are falsified in others, which eventually leads to the children constructing culturally specific theories of mind. However, there is no qualitatively new mental content possible – the cultural concepts are merely recombinations of the universal foundations. The Welsh concept of *hiraeth*, for instance, which means a specific kind of homesickness (Polk, 1982), and which Welsh people report to feel and undoubtedly read others as feeling, will have to consist of some highly nuanced mixture of desire and other purportedly innate concepts. Although in theory such constructivism could exist, it is highly controversial to take it to be an adequate model of how humans develop cognitively: Remember, it implies that the truth-bearing properties of the mental structure all come from the foundation, and so must even such concepts as car sickness or e-mail.[5] Consequently, even though a nativist may claim that "novel concepts can be acquired, and new principles of attribution learned, relying both on individual experiences and cultural input" (Westra & Carruthers, 2017, p. 166), it seems that there is little to the effect of an actual model of how that might be so. With the constraints intrinsic to the ToM framework, enculturation can be nothing more than reconfiguration of innate contents.

2.3. The frame problems and holism of mental-state attribution

From a more epistemological perspective, encodingism faces problems when the context sensitivity of natural cognition is concerned, which has been discussed in AI literature under the label of the frame problems (Ford & Pylyshyn, 1996). Let us first discuss briefly the general nature of the frame problems, after which we will turn to their expression within the context of social cognition.

The underlying presupposition of encodingist frameworks is that explicit representation is the most fundamental kind; it is by encoding information about objects, their properties and relations, that an agent manages to successfully interact with reality, to successfully anticipate what will happen and act accordingly.[6] However, we behave in agreement with an unbounded set of aspects of reality without representing them explicitly, and so do other animals (e.g. all organisms behave in accordance with gravity, but only humans have managed to represent it explicitly) and so this presupposition is plainly wrong. If we were to function in such explicit ways, the tasks would be computationally and memory-wise impossible: Any aspect of the situation, and past situations, can be potentially relevant for the present interaction, and so the system would have to encode an unbounded set of relevancy relationships and compute their applicability for the present context every step of the way. This is evidently an impossible undertaking (for a more thorough discussion, see Bickhard & Terveen, 1995, pp. 214–233).

Thus, any purported explicit representations or encodings that are claimed to drive the agent under determine their actual behavior, and consequently – what is most relevant for the present discussion – any *attribution* of such explicit representation to another agent underdetermines prediction of their behavior. The latter has been pointed out in social cognition literature (Morton, 1996; Zawidzki, 2013). As these researchers note, mindreading models cannot account for competent social cognition due to the holism of propositional attributions: Mental state attribution is a poor predictor of another's behavior because how a person acts is determined by an unbounded set of possible mental state combinations and other parts and aspects of their cognitive system and situation. Beliefs and other mental states attributed by the purported ToM mechanism translate into actions only together with the context of the whole of the organism's state – they do not determine what a person will do. I might believe that it is raining, but still not take out my umbrella if I wish to get wet, for instance. What is more, the holism problem becomes even greater as mental states relevant to future behavior can be culturally specific, the acquisition of which is, as we have argued, a more fundamental problem on its own.

Zawidzki (2013) claims that the problem of holism is solved by processes of mindshaping: Humans living in a society become alike in how they think and behave – their culture shapes their minds – so that attribution of mental states becomes tractable. Mindshaping certainly takes place, and our proposal provides a model of it too, but we believe that the problem of holism goes deeper than Zawidzki considers. He remains explicitly agnostic about the mechanism underlying social competence, staying open to an encodingist mental-state attribution mechanism doing the job (2013, xiv-xv). Our claim is therefore stronger in that we believe the frame problem persists as long as we stick to an encodingist framework; a cognitive system cannot possibly be driven by explicit encodings to begin with, which means that no attribution of such encodings can successfully predict its behavior, no matter the amount of mindshaping that has taken place.

Some cognitivist accounts have attempted to tackle this problem – e.g. Apperly and Butterfill's "tracking" beliefs through representation of "belief-like" states (2013) – but the fact remains that encodingism does not provide resources to model competent interaction without explicit representation of the thing interacted with; it does not model implicitness, which clearly permeates all of cognition (we do not "token" all the beliefs we "hold" about reality and encodingism requires that we would).

2.4. The coordination problem

Holism is closely related to another frame-problem issue – that of social coordination (Schelling, 1997[1963]). Briefly, in order to correctly anticipate behavior in most *interpersonal* situations, one needs to consider that another's behavior is contingent on how they represent *my anticipation of their behavior*, which fact, in turn, has to be taken into consideration by my anticipation, leading an infinite regress of iterations. Hence, the intractability of holism is aggravated even further when we consider actual interpersonal interaction, not just third-person observation of another's behavior. The hierarchy of reflexivities in the ontology and epistemology of social situations is intrinsically unbounded, and the encodingist commitment to explicitness requires that

[4] Note that distributed processing, such as in connectionism or dynamic systems, which are sometimes claimed to be an alternative to classical symbol architectures can be just as encodingist as their predecessor. Dynamic systems is more nuanced, however, as it does not presuppose an encodingist metaphysics and can be given a more tenable metaphysical interpretation (such as the one discussed in the present text). Classical symbol computationalism and connectionism do presuppose an encodingist metaphysics and thus are untenable at the outset. For a fuller criticism see Bickhard (2015b); Bickhard and Terveen (1995).

[5] This theoretical necessity is perhaps best illustrated by Fodor's radical concept nativism (Fodor, 1975, 1981).

[6] By "explicit" we mean roughly "actually present in the content of the representation", not "conscious" as it is often understood in the literature.

the entire hierarchy be explicit — and that will not fit into a head (or solar system, or the universe).

It is interesting that ToM research almost entirely ignores the problem. The debate has for the most part presupposed that mental-state attribution does the trick, and focused instead on how the capacity for it is implemented and acquired developmentally. However, mind-reading is an impossible prediction tool – its purported main function – even from an observer's perspective, and even less so in interpersonal situations that require reciprocal characterization between the interactants.

We believe that this issue, and the ones discussed previously, are solved once we shift to an action-based model of representation, one that models representation on the relation of implicit presupposition, rather than explicit encoding, as we discuss in the latter part of this paper. The shift, we believe, not only solves the problems, but also allows for a more fruitful model of socio-cognitive development and its enculturation. We present the positive model below.

3. Interactivism – basics

In this section we present the framework of interactivism, which we believe avoids the problems discussed above. This section is concerned with metaphysical foundations, while the next one will focus specifically on the interactivist model of social cognition and its enculturation.

Interactivism falls within the wider family of approaches that adopt an action-based ontology of the mind. The action-based conception derives in varying degrees from a number of theoretical predecessors – Piaget, Vygotsky, American Pragmatists, Gibson's ecological theory of perception, or later Wittgenstein's model of language – and its other most notable models today include Carpendale and Lewis' social-constructivist approach (Carpendale & Lewis, 2006, 2015) and Nelson's Community of Minds (Nelson, 2007). These approaches are the closest to interactivism, and it might be viewed as a low-level complementation of C&L's and Nelson's more theoretically lean models. Of slightly different provenance (much greater influence of phenomenology), but still largely convergent with the action-based conception of cognition, are enactivist approaches (see the review in Ward, Silverman, & Villalobos, 2017). Here disagreements are greater, but it is clear that they too are pointing in the generally same direction as the interactivist model. We believe the unifying thread of action-based approaches to be the claim of ontological emergence of the mind through or within (social) interaction, and the constitution of cognition by interactive anticipations rather than encoded information. The proposed model's affinity to the above proposals should become apparent in our exposition below, but the interested reader might consult previous work that has explicitly compared some of these and other accounts with interactivism (enactivism – Bickhard, 2016a; predictive processing – Bickhard, 2015c, 2016b; Mirski, Bickhard, Eck, & Gut, 2020; an exhaustive review of various models – Bickhard & Terveen, 1995).

One of the aims of interactivism lies in making explicit the foundational presuppositions of the action-based conception, solving any conceptual issues that might emerge thereby, and proposing a detailed account of the most fundamental aspects of cognition. This inevitably involves a substantial amount of explication of other frameworks and introduction of new technical concepts, some of which would perhaps not be entirely acceptable by other action-based theoreticians. One central such issue is the interactivist retention of the concept of representation, which many action-based approaches either avoid or actively attack (e.g. Hutto & Myin, 2013, 2017). We believe that this animosity to the concept stems from unwarranted equating of "representation" with its encodingist model, which – as discussed in the previous section – we also claim to be untenable. "Representation" can, however, be given an action-based interpretation, one which is unlike the encodingist standard and which can be highly illuminating for an action-based theorist, especially in the context of development and enculturation. The interactivist concept of representation or representing comes about naturally from an analysis of the dynamics of cognitive systems and the consideration of their metaphysics – it is not arbitrarily and normatively projected on it as is the case with the encodingist model that borrows the concept from artificial intelligence and attempts to force it onto natural systems. The core problem of representation is and has always been that of accounting for truth value – that organisms can be wrong or right about what they 'assume' reality to be is a problem that needs explanation. The basic claim against encodingism has at its core that encodingism cannot account for truth value, which is why this has been a core problem since the Pre-Socratics. Interactivism claims to account for truth value, and thus claims to account for the core problem of representation/representing. Similar treatment is given to some other extant concepts – such as "concept", "intension" or "mental content" – instead of being abandoned altogether they are explicated in terms of the framework. Sometimes this results in a diverse ontology proposed in place of what has been traditionally assumed to be a unitary phenomenon; this is the case for the notion of a concept that in its traditional meaning spans what interactivism identifies to be a fairly diverse dynamics (Bickhard, in preparation). In short, the general strategy adopted in interactivism is to treat the extant conceptual landscape as needing a new, tenable interpretation rather than as inherently doomed.

Interactivism has been formulated with the use of various formalisms – such as automata theory (Bickhard, 1980), dynamic systems (Bickhard, 2015c; 2015d), or topology (Bickhard & Campbell, 1996) – and more formal treatments of what is discussed here are naturally possible. However, the main concern of the current paper lies not in the adequacy of any particular formalization, but rather in fundamental metaphysics of cognition and social reality: the discussion concerns the fundamental nature of what the various formalisms can be used to describe, but which cannot be fully captured by the meanings inherent in the languages of any currently available formal system. For that reason, although necessarily technical, the following discussion is presented in informal terms.

3.1. Representation

Mental representation is a kind of normativity. Thus, the problem of normativity is a more fundamental one, and the question of how normative phenomena can exist in a world of causes has to be answered. Interactivism models normativity as an emergent property in far-from-thermodynamic-equilibrium (FFE) dynamics (R. J. Campbell, 2015). FFE systems by their very nature need maintenance if they are to persist – they need some processes (either internal or external to the system) to provide the necessary conditions for their continued existence (persistence). That a certain process can contribute or not to the maintenance of an FFE system constitutes those processes as functional or dysfunctional relative to that continued existence, which gives us naturalized normativity (Christensen & Bickhard, 2002).[7]

Living organisms are a special kind of FFE systems: They are recursively self-maintenant. This means that they not only constitute part of the processes ensuring their own persistence (self-maintenance), but also that they can switch how they operate in different environments to such functional organizations that serve the overarching function of self-maintenance (recursive self-maintenance). When considered as processes, living things can be said to be flowing, differently in different

[7] *Serving* a function is primary for us. *Having* a function is derivative from serving a function; a process has a function if it is functionally presupposed in some capacity (as serving the function) by another process. The present *synchronic* account of normativity of function is in contrast to etiological models of function, which see *having* a function as primary and define it diachronically (Dretske, 1988; Millikan, 1984, 1995). We believe the etiological proposals to be insufficient; function normativity is claimed to be constituted by the history of selection, which renders it synchronically causally epiphenomenal (Bickhard, 2015).

conditions, so that they do not dissipate into equilibrium, and the organization of the functional processes that constitute that flow is the subject matter of life and social sciences.

To stay self-maintenant across a number of different environments, an organism needs to be able to detect those environments, and to switch to such a mode of functioning that will enhance its probability of survival in it. The former needs to be distinguished from the latter; while it is necessary for an organism to be able to detect environments, to come into epistemic contact with them, it is the adopted mode of functioning that is either functional or dysfunctional, and thus exhibits normativity. When an organism adopts a certain way of functioning, it implicitly presupposes the environmental conditions that are necessary for the mode to be functional, to actually contribute to the self-maintenance of the system. When *E. coli* detects that it is moving down a sugar gradient, it tumbles – when it is moving up, it swims. In doing so, it implicitly presupposes that the environmental conditions are such that the adopted mode of functioning is self-maintenant – it presupposes a certain orientation to a distribution of sugar concentration. The truth of this presupposition matters a great deal for the organism because its FFE existence depends on it.

Given the complexity of the environment in which complex organisms such as humans stay self-maintenant, it makes sense that they do not automatically react to detected conditions with some fixed mode of functioning, but rather divide their internal organization into two basic kinds: one that indicates what interactions are possible, and one that selects what interaction to actually engage in, given the present state of the organism. The former are the processes that have the *function of representation* for the organism, and exhibit causally-efficacious semanticity – they are used by the organism to "probe" reality for possibilities of interaction.[8] This allows the organism not to be triggered by a stimulus, but to consider a wider spectrum of interactive possibilities in the represented environment and settle on the one that best serves its current goals.

Interactivist representation describes the organism utilizing the normativity of its FFE subprocesses for the fulfillment of its goals. The organism's interaction with the environment has an aspect internal to the organism; that is, interaction is accompanied by an internal anticipatory (FFE) process, which implicitly presupposes conditions necessary for its persistence. Representational processes of the organism use that property of implicit presupposition in order to indicate what is *possible* in the given situation, defining the situation as such where some set of interactions can be potentially engaged in. This can be done because the world exhibits regularities that determine contingencies between possibilities of interaction, and these regularities are informationally redundant – a presence of a tiger in the vicinity, for instance, can be accessed in many ways, such as spotting its tracks, other animals running away, or hearing its roar. This means that the organism can use interactions to differentiate environments and indicate what other interactions are possible. For example, imagine there were two kinds of fruit in the habitat of some organism – one firm and the other soft. The firm one is sweet and edible, while the soft one bitter and nauseating. The organism could then use squeezing the fruit as a differentiating procedure, to differentiate between situations that afford eating, and those that do not. For that, all that is needed is an internal organization of processes that makes (the internal aspect of) eating conditional on how (the internal aspect of) squeezing flows. Interactive indication is the interactivist model of (logical or functional, not necessarily linguistic) predication and hence representation.

[8] The concept of "interactive indication" is similar to Gibson's concept of affordance. That is, "interactive indication" is very close to one of the number of readings of "affordance" that can be found in the literature, the one that identifies it with the internal processes "readying" the organism to do what the perceived environment affords (Rucińska, 2020). For a detailed explication of Gibson's theory in interactivist terms, see Bickhard and Richie (1983).

The process of setting up of the organism's anticipatory processes is termed *microgenesis*, and involves the totality of interactions indicated as possible in the current situation. It needs to be stressed that microgenesis does not explicitly represent the conditions necessary for the indicated interaction to succeed. It does not encode anything. Rather, it implicitly or functionally presupposes the necessary conditions of its stable flow, and these implicit presuppositions are what constitutes mental content within the model. The relation of presupposition, unlike that of encoding, is a naturalistically tenable basis of mental representation (i.e. it follows from the metaphysical considerations of FFE dynamics of living things). [9] Relevant to the present discussion, implicit presupposition does away with the frame problems in general, and with the problem of holism and infinite iteration of mutual characterization in the coordination problem in particular: Compatibility with the anticipated flow is *implicitly presupposed*, not explicitly represented, and thus does not impose an impossible computational and memory load. We discuss this in more detail in later sections.

3.2. Learning and development

The internal organization of anticipatory processes can be the product of phylogenetic selections and develop biologically in the organism, without the involvement of experience with the external world. It is plain, however, that complex organisms acquire it in ontogeny via their interaction with the world. Hence, a model of learning and development is needed. Unlike encodingist models, which rely on explicit representation, the interactivist model does not necessitate prescience about what is to be learnt – no foundationalism. An organism capable of learning needs to be capable of generating new process flows – in the limiting case, by random variation – and "checking" if they stabilize. In plainer language, the organism needs to be able to try different interactions, and check if they succeed. Then, assuming it is possible for *success* to be retained as a permanent change in the functional organization of the organism and *failure* to be selected against, we have a basic mechanism of learning – variation and selection learning, or evolutionary epistemology (D. T. Campbell, 1974).[10] It should be easy to envisage that, given a certain level of complexity, the organism will be able to create a vast web of iterated interaction possibilities by interacting with its environment and trying out what it can do in it. Importantly, this will be a (recursively) constructive process – once some organizations are retained as a result of successful interaction, they will form the context and ground for further interactions, often opening up possibilities for success at interactions simply not possible at earlier stages.

A useful way of looking at the above dynamics is with the concepts of situation knowledge and world knowledge; we will make use of them when discussing social cognition. Situation knowledge is the present state of the organism's indications of interactive possibilities. World knowledge is the totality of knowledge inherent to the system's organization of how to maintain, update, and modify situation knowledge, on the basis of acting and perceiving. In other words, world knowledge is the inherently hypothetical or conditional knowledge of all possible interactions that the organism is capable of, and situation knowledge is

[9] This constitutes a shift of the metaphysics of representation: away from (encoded) Correspondence to: that which bears truth value", and it is future oriented anticipations or indications that can be true or false — that bear truth value.

[10] Anticipatory success/failure is a property of processes of recursive self-maintenance of a FFE system. Those processes either contribute (success) or not (failure) to the maintenance of its ability to self-maintain itself. The functionality of such processes lies in that they involve implicit anticipations (which can be true or false) about properties of the environment that are relevant for how the system should reorganize itself to stay in a viable relationship to its FFE conditions.

the knowledge of which of those interactions are (anticipated to be) possible at the moment. For instance, I can know how to ride a bike in general (world knowledge), but to know that I can ride a bike in the present situation some organism-internal conditions need to be fulfilled, such that my world-knowledge structure of "bike riding" is indicated as *currently* available (e.g. certain sensory processes involved in seeing a bike or the bike mechanic telling me that my bike is ready to be picked up – what indicates a structure is a matter of learning). When those situation knowledge conditions are fulfilled and some interactive structure is indicated as currently available, we say that the organism *apperceives* that structure (not as such, that is, but implicitly – it apperceives the potential process flow that constitutes that structure).

Interactive structures of an agent's world knowledge constitute the world for the agent. The interactive regularities in the world are fairly general – not every bike is the same in all aspects – but bikes do, on the whole, exhibit interactive commonalities: e.g. they can be touched or scanned visually in predictable patterns. It is such interaction-based general schemes for objects and other interactive patterns in the world that will eventually constitute the child's world knowledge.[11] These interactive patterns will constitute the primary representational currency of the interactivist agent's situation knowledge; they will constitute substructures of the world knowledge and will be indicated as accessible in the situation knowledge. Learning and cognitive development will thus start off on a micro-scale as variations and selections in minute aspects of the anticipatory flow (e.g. muscle activity while manually manipulating an object or anticipating its behavior in vision, cf. Piaget, 1954), but once such interactions are mastered, and the structures responsible for them are part of the child's world knowledge, development will proceed at a more general level where variation will be mainly between a number of already extant structures (e.g. trying to open a box, the child might attempt to smash it against the floor, pry it open with some other object, or look for the key; she will not start the learning process entirely *de novo* as the system is already organized in some relatively stable manner as a result of previous variation and selection learning). Development is therefore a constructive process in which simple interactions need to be mastered first before more complex ones even become a possibility, resulting in an increasingly more complex world knowledge of the child and her growing competence as an agent.

3.3. Reflection

As has been mentioned, the action-based knowledge discussed above is entirely implicit – it does not represent its own correctness conditions (implicit presuppositions). Roughly, it is the interactivist model of what in contemporary literature is studied under the label of embodied knowledge. Interactivism, however, does offer a model of explicit thought, something which is generally problematic in embodiment frameworks (see, e.g., De Bruin & Kästner, 2012; Edelman, 2003; Goldinger, Papesh, Barnhart, Hansen, & Hout, 2016).

With experience, the functional organization of microgenetic processes (the world knowledge) will form a relatively stable organization. This organization will exhibit properties in its own right, which can be known themselves (they will *instantiate* concepts for the organism). Interactivism models explicit thought as a second interactive system that anticipatorily interacts with level-1 process organization and thus can represent organizations of its implicit presuppositions explicitly.

The crucial point to note is the limitations of basic knowing. Level-1 anticipations are anticipations of *realms of potential internal process flows*, they are not anticipations of things in reality, not even anticipations of some "contentless" discrete mental units that correlate with features of reality. The world of objects, events, and people is not what a level-1 knower experiences.

Level 1 does, however, instantiate or embody the furnishings of our reflective mind. The idea here is that the objects of reflection – items, people, events, types of change and interaction, etc. – are constituted by what will support functional patterns at level 1. A toy, for instance, generally affords the same stable interactive possibilities – it can be manipulated, scanned visually, thrown, given to somebody, etc. – and can be returned to later from some other place. Reflection anticipatorily interacts with these patterns and abstracts their properties, forming explicit representations. This allows the organism to consider different interactions "offline" as well as greatly improve interaction online by implementation of explicit learning strategies, rehearsal, planning, inhibition, and other executive functioning.,[12][13]

4. Interactivism as a framework for social cognition

The above has for the most part addressed the problems we pointed out in our earlier criticism. The anticipatory, action-based model of knowledge together with evolutionary epistemology solves the copy problem and foundationalism, as well as the frame problems: (1) Anticipatory processes can be in the limiting case generated entirely at random and retained if and after they succeed, which means that neither independent epistemic access to reality nor foundational content are needed in this model of learning (Bickhard, 2003, 2010). (2) The frame problems are not an issue due to the fact that content is constituted as implicit presuppositions of anticipatory processes (Bickhard, 2001). These presuppositions are unbounded, and while it might be useful for the organism to explicitly represent some of them in reflection, such explicitness is not necessary for the organism to interact competently with its environment.

It also becomes clear that there are no in-principle obstacles for the model to account for the enculturation of (social) cognition, but a more thorough discussion of this point is due. Below, we present a (necessarily compact) model of socio-cognitive development within the interactivist framework (Bickhard, 1992a, 2008, in preparation, 1980; Christopher & Bickhard, 2007). We hope to show that interactivism naturally accommodates the fact of socio-cultural modulation of socio-cognitive abilities, while at the same time remaining ontologically tenable.

4.1. Situation conventions

From an interactive perspective, knowing is anticipating interactive possibilities. Thus, social cognition is knowing how to interact with other agents, how to anticipate their behavior given one's own actions. This is in-principle unproblematic (though potentially complex) when we consider interacting with agents that do not learn (e.g. a pre-programmed robot): All there is to learn in order to represent such an agent is to know how they react to one's actions. Although more complex and involving a much greater context-sensitivity, this can be achieved in ways analogous to how one interacts with physical objects. However, given that even very simple organisms can learn, social cognition is usually trickier than that. Agents that learn acquire their behavioral dispositions during ontogeny (i.e. they construct their world knowledge), and thus their learning histories have to be accounted for in anticipating their behavior. This is a problem not only because it seems

[11] For a more thorough presentation of schemes and other functional structures within the framework see Bickhard (1980).

[12] Importantly, level 2 will also exhibit properties, which can be known by a third interactive level, which will also instantiate some properties, also potentially known by a one level higher reflection. And so on. See R. L. Campbell and Bickhard (1986); Bickhard (1998).

[13] In this presentation, we have skipped the interactivist model of emotion. The reason for that is mainly the limited space. We plan to extend the proposal with a discussion of emotion in further work and for now can only refer the reader to Bickhard and colleagues' work on the subject (e.g. Bickhard, 2000; Bickhard & Campbell, 1996).

to necessitate learning of the interactive potential of every single individual with their unique history, but most significantly because the individuals we interact with, in turn, learn how to anticipate us as well. In an interpersonal situation, I have to account not only for my interactant's representation of the physical reality – e.g. what objects they see and how they view them – but also for their representation of my own interactive potential. Since my interactive potential clearly hinges on my representation of their interactive potential, this seems to lead to an infinite regress of reciprocal characterization – the coordination problem we mentioned before.

D. K. Lewis (2002[1969]) proposed that the coordination problem is solved by convention, and Bickhard (1980) adopted and adapted the idea into the interactivist model. The basic claim is that we *do not* achieve coordination by mindreading (i.e. encoding information about) another's mind, but by presupposing that they will behave in the way consistent with how we coordinated actions before in a given context.[14] Although it might seem like another vicious circle, it is not – the original coordination success can be achieved by random variation – both parties stumbling upon mutually complementary anticipatory characterizations of each other (the chances of that occurring are clearly aided by innate scaffolds in humans).[15] Once retained (likely after a number of occurrences), a given way of coordinating action in some situation will be the anticipated default for both individuals. Such commonality in the characterization of a situation is a *situation convention* and it forms the basis for social cognition and social reality in interactivism (Bickhard, 1980, 2008). Situation convention, in essence, is a socially emergent "anticipatory agreement" concerning what to do in a social situation, how to interpret each other's actions given the context we are in.[16]

Consider an example (simplified to be illustrative). There are many things two individuals living together can do first thing in the morning. Among others, they can wander, go hunting, or sit down and eat. Depending on what an individual decides to do – what his selections within his web of situation knowledge are – anticipation of his behavior by the other person will differ: His situation knowledge must be organized in a way anticipated by me for coordination to be possible; and if he is to interact with me, he has to do the same about my situation knowledge. While there is no way to know in advance what the other will decide to do in the morning, it is in principle possible in the present model to stumble upon a mutually consistent characterization of each other's situation knowledges, which will result in an anticipatory success of both parties and will set precedence for retaining the functional structures responsible for the successful anticipation. If we both happen to assume that we both eat first thing in the morning,[17] and this is indeed what we are doing, then we set precedent for future anticipation. Thus, a situation convention will emerge about what one does in the morning, and I will anticipate the other person to behave in accordance with it. Notice that even though the above example is rather coarse-grained, the same principle can account for mutually consistent interpretation of gestures, or vocalizations, and in fact any other form of interpersonal activity: They all need to be interpreted in a consistent way in order to lead to coordinated anticipations among the individuals involved in the social situation – they need to be apperceived as situation convention.

The above case applies to characterizations of particular individuals in some particular contexts; that is, the situation convention is entirely contingent on past interactions between the individuals involved within some specific settings. As such, it is likely to exhibit a great deal of idiosyncrasy resulting from the particularities of the ways coordination was achieved between the interactants, and not work with another individual who did not take part in those past interactions. This is clearly the case in early childhood family contexts, where the parent and child often develop peculiar communication standards (e.g. Carpendale & Carpendale, 2010).

Naturally, we interact with many individuals, and so we need to be able to know what situation conventions are available with them. One way of achieving this would be to differentiate conventions on the basis of individuals and our corresponding interaction histories with them. It is evident developmentally as newborns already differentiate between persons (Bushnell, 2001). And it is certainly the case in adulthood as well, as we all have conventions specific to individual people we know personally; it is to a certain extent what we mean when we claim that we know each other. Notwithstanding, a much more effective way of coordinating behavior on a wide social scale is by *institutionalization* of situation conventions. First, however, let us consider *typification* of situation conventions, which will make the presentation easier. A convention is typified when some cue (part of my situation knowledge) is anticipated to make another person enter or at least apperceive the situation convention. Typification can be one-sided: the caregiver might anticipate that the child will apperceive a toy as an indication of the situation convention of joint play, but the child herself does not have to anticipate the same about the caregiver (she only knows the situation convention of play, not that her caregiver will apperceive it when presented with a toy). Once she does, however, the typification becomes mutual and the toy can be used as a means to evoke the situation convention of joint play. Such mutually typified conventions that can be invoked by some means are termed institutionalized conventions – it is presupposed that we both know what kind of situation the toy indicates. Institutionalized conventions will tend to spread in a society as people bring conventional structures acquired previously in interaction with other people to every new social situation. They will thus attempt to solve new coordination problems in ways that worked with other people (as the relevant schemes are already in their world knowledge), setting precedent for the establishment of the same type of convention with a new person. As a result of this process, it will become a taken for granted fact that certain objects or behaviors indicate particular types of situation conventions and the need for conventions uniquely suited to every

[14] Remember that implicit presupposition does not necessitate encoding the infinite number of reciprocal characterizations.

[15] We are aware of the problematic nature of the concept "innate" given the multiply contingent dynamics of development (Oyama, 1985/2000; Samuels, 2004). Here we have decided to use it when talking about a particular kind of scaffolds. With "innate scaffolds" we wish to express that certain organismic forms are a reliable developmental outcome for a species (they are not "preformed" before development occurs, as has been assumed by core knowledge accounts, e.g. Spelke & Kinzler, 2007, but rather emerge reliably in a species across its usual environments) and that these forms are a resource for the process of microgenesis – they constrain the space of anticipatory "guesses" that an agent has available in an interpersonal situation. We acknowledge that that does not do away with the problems with the concept, but we felt the need to differentiate such scaffolds from the ones that might come from the environment as well as the ones that are constructed via microgenesis.

[16] It needs to be stressed here that the meaning of "convention" stipulated here is both more specific and more general than the ordinary meaning of the term. It is more specific in that it refers to a certain class of anticipatory processes, namely such that are in relation of mutual anticipatory consistency. And it is more general in that the class will cut across other classes of processes, transcending its ordinary-language restriction to generic cultural phenomena (e.g. animal alarm calls are by the present definition a case of convention but would hardly classify as such in the ordinary language use of the term).

[17] This example nicely illustrates the role of innate scaffolds for the construction of situation conventions: Organisms need sustenance in the morning and so eating is what they usually will do. The main point is that embodiment and structures selected for in phylogeny will narrow down the space of potential anticipatory "guesses", leading to more effective success at coordination and establishment of situation conventions. It is important to note, however, that this is not *necessary* in the present model – variation and selection process *can* be entirely blind in theory. This contrasts with encodingism, which necessarily requires inborn concepts in order to get off the ground.

single individual will be avoided.[18]

Notice that the process of institutionalization of situation conventions inherently involves conventionalization of the ways in which situation conventions are invoked in social interaction: Building on the above example, joint play can be indicated not only by a toy but by any arbitrarily chosen aspect of the situation knowledge. Any situation convention is constituted by commonality of characterization of the situation in the situation knowledges of the interacting agents; it is not *necessarily* constrained by any other (non-conventional) part of the situation. In other words, situation conventions can be actualized solely by the interacting individuals as long as they manage to actually characterize the social situation as the same kind of social situation. Due to the inherent arbitrariness, the ways of identifying and invoking situation conventions in social situations will necessarily be conventional themselves, as all interactants need to apperceive them as establishing the same situation convention. In effect, different gestures, sounds, or elements of the context will emerge in a society as conventional indications and transformations of situation conventions.

Such conventions for invoking and modifying conventions are the basis for interpersonal communication, a foundation of language. In essence, they will function as manipulation of the interpersonal situation: When a child brings a book to an adult, for instance, she attempts to transition to the book reading convention, and her gesture is a conventional way of effecting that change. Language, within the present framework, is of the same kind, but its productivity and fine-grainedness allows for much more nuanced situation convention manipulation; we discuss language in section 4.3.

4.2. Social reality

Let us now take an eagle-eye perspective onto the proposed model of social cognition. The main claim is that social reality is constituted by "the factual structures of relationships among world images [or knowledges] that constitute institutionalized conventions and the factual structures of relationships among those conventions" (Bickhard, 1980, p. 78). Social reality is therefore supra-individual functional organization emergent in the dynamics obtaining within and between society's individual minds, as a consequence of their attempts to coordinate their actions in whatever interaction realm they happen to be interacting.

It is important to appreciate the nature of social reality conceived thus, and a comparison with the representation of physical reality is illustrative here. When an individual learns to interact with the physical world, she constructs functional structures that correctly anticipate possible interactions with the elements of the physical world. An object (psychologically) from an interactivist perspective is a stable, recoverable functional organization that is embedded in the wider functional organization of the world knowledge. Thus, an experienced agent's world knowledge of the physical world will be constituted as factual structures of relationships within and between such recoverable substructures. Now, the child's development of her social cognition is akin to that, but the resultant functional organization that represents social reality at the same time participates in the constitution of that reality. Social reality "is a world that originates in [people's] thoughts and actions, and is maintained as real by these" (Berger & Luckmann, 1967, p. 32). Thus, interactivism offers a naturalized model of social reality, including institutions and intentional objects, which has been a major challenge in social sciences (e.g. Eck, 2015; Legros & Cislaghi, 2020; Turner, 1994, 2018).

Importantly, the model accounts for the normativity of social reality. Since culture is constituted as factual relationships between world knowledges of its participants, its normativity is derived from the normativity of the FFE dynamics of the individual minds. It clearly goes beyond an individual mind, however, as an individual's anticipations can be wrong with respect to the dominant social reality that inheres in the minds of other people, and conventions, once established, will exhibit a great deal of independence and self-organization. Thus, the supra-individuality of conventions helps us understand the independence of cultural norms from the individual. It also helps us understand the inner conflict that sometimes occurs between the conventional normativities instantiated by an individual mind, and its other, more personal normativities. Moreover, the framework makes sense of the truth value of statements about intentional objects, such as Little Red Riding Hood – it is constituted by the correctness of the anticipatory presuppositions about relevant structures of social reality: e.g. if my anticipatory schema for the story involves representation of a fox instead of a wolf, it will disagree with the convention, which fact might be discovered if interpersonal interaction with other people or physical elements of culture (e.g. a book) fails to support my anticipatory flow that presupposes the fox being involved instead of a wolf.

Social reality is emergent — starts off as a solution to coordination challenges in joint interaction with physical reality. However, once some conventions become established in society, they become a realm of interaction themselves, and as such can lead to further coordination challenges that are embedded *within* conventional reality. And once the solution to those challenges becomes also part of the extant social reality, some yet newer coordination challenges might emerge, for which further conventions may develop; and so on, leading to a social reality gradually more removed from the "natural" world. Clearly, this ontological climb does not proceed forever, and there are other factors that modify social reality as well.[19] These issues are complex and we will not discuss them here; the crucial point for the present discussion is that conventionalization of minds – and thus the emergence of culture and social reality – is a constructive process both ontologically and historically (and phylogenetically)[20] with higher-level conventions emerging in the context of the existing ones, often building on, presupposing, and likely modifying prior social and institutional organization.

Further, the framework has important consequences for the ontology of the person. In interactivism, a process or a system is individuated on the basis of its interactive properties – a thing is how it interacts (potentially) with other things. It becomes clear, then, that a considerable, perhaps most significant, part of a person is constituted by convention; it is shaped by the culture the person has developed in (cf. Zawidzki, 2013), as it is through their participation in or co-constitution

[18] The differentiation between representations of specific persons and conventionalized roles has been explored recently by others (e.g. Fiebich & Coltheart, 2015; Newen, 2015). The present proposal provides a wider framework that accommodates these remarks.

[19] An emergence of a group of people who can specialize in acting within social reality might play an important role. Such an emergence can be enabled by many factors, such as growing population or overabundance of resources. Another clear factor is informational constraints: a society with a writing system can clearly establish and maintain a much wider and more complex convention.

[20] The framework is fully consistent with the possibility of conventions that are not the product of experience with other agents but develop via dynamics not involving inter-agent interaction. Such "innate" conventions undoubtedly exist – e.g. alarm calls in animals, or a newborn's crying. They are convention because they are apperceived as indicating the same situation for every member of the species, but they do not require past interaction and coordination with other agents to be apperceived as the same conventions. Such conventions are arguably one of the most significant innate scaffolds with which cognitively constructed conventions are build. Heyes (2018) proposes a somewhat a similar idea, but with important differences.

of convention that they are disposed to interact in a particular way with their surrounding reality, including other people.[21] Indeed, the proposal provides a naturalized framework to understand the claim that "there is no such thing as a human nature independent of culture" (Geertz, 1973, p. 49).

Finally, let us frame the above considerations in the form that is more in line with how social cognition is investigated in ToM debates: What does a person represent when they apperceive another person? Our answer to that is that their representation is thoroughly conventional as most of the possible interactions with other people are constituted by conventions (as are those people). Granted, the conventions will involve non-conventional structures such as of physical objects or the interactant's body, and they will form in accordance with all kinds of biological or physical constraints, such as the informational channels of the human senses. However, these will be aspects of situation conventions, and will not be – at least initially – cognitively differentiated from them (they will be *implicit* in the conventional structures). Minds that have grown in a society are inherently cultural, and any biological human universals that are discovered are necessarily abstractions from something that is in its current form a result of complex influences involving biology as much as culture (Oyama, 1985/2000). In contrast to the essentialist assumption of the ToM models, there is no default, culture-free, theory of mind module to which cultural concepts are added, but rather minds that are organically encultured and that view other minds in their encultured ways.

4.3. Language

Language is an interaction system (a part of the world knowledge) specialized for manipulation of situation convention. In its essence, language is highly precise and systematic convention for invoking situation conventions that achieves its goal via intermediate manipulations of the situation convention (Bickhard, 1980, 2007). For instance, the utterance "The toy is broken" involves "the toy" modifying the participating situation knowledges such that the next modification – "be broken" – will apply to that part of the situation knowledges that represents the toy and will modify it accordingly to what "be broken" means.

Development of social abilities will consist in the child's exploration of possible manipulation within situation conventions, peaking in the development of a language system that will allow the child to selectively target highly specific aspects of situation conventions and thus differentiate those aspects as cognitive (interactive) units. Although verbal behavior accompanies even the earliest conventions, the central property of the language system – its productivity – can only emerge gradually, by increasing differentiation of situation conventions. For instance, the child might first discover that pointing transforms the situation convention into one that involves the object pointed to. She will naturally experiment with it and discover its great potential across different social situations. In the process of this experimentation, the child might discover that uttering a sound while pointing has a systematic effect onto how the situation convention is transformed. It will do away with much of the ambiguity involved in pointing itself, and will become a more reliable transformation tool (Colonnesi, Stams, Koster, & Noom, 2010). Then, the child might learn that adding another word after the noun will specify the transformation even further.[22] Clearly, it is not long before the child discovers the systematic functionality of words and their combinations, at which point her engagement with language will increase drastically – which indeed seems to be the case in the second year of life (McMurray, 2007). This is naturally hardly a full account of language acquisition, but it should illustrate the principle of language's ontogenetic emergence within the present framework.

A few words are due about how language – or rather situation conventions or meanings it manipulates – represents. Most fundamentally, as has been already discussed, situation knowledge represents via indications of possible interactions (world knowledge structures) that functionally presuppose certain features of reality. The structures of situation conventions, in turn, represent via functional relationships with the structures that indicate such real-world interactive possibilities. Early and simple cases of this functional link will be highly coarse-grained, such as a gesture for handing one an object. However, with the fine-grainedness afforded by language, the situation convention can be differentiated into the part that specifies the object – e.g. the toy, and the part that specifies what about the toy's representation is to be manipulated. And analogously for all the other linguistic utterances and their corresponding functional effects onto the situation convention. Moreover, once the child is capable of reflection, the functional organization structured by linguistic use can be represented explicitly, which explains why it is easier to think about things that have names for them.

Naturally, the above is an extremely brief presentation; it should suffice, however, for our present purposes. For more detailed breakdown of issues directly related to (psycho)linguistics see Bickhard (1980, 2007, 2015a), and Bickhard (2008). See also Kempson, Cann, Gregoromichelaki, & Chatzikyriakidis, (2016, 2017) for a convergent approach to language.

4.4. Some comparisons with enactivism

Considerable similarities can be discerned between the interactivist model of language (and social reality more broadly) and the recently proposed enactivist alternative (Cuffari, Di Paolo, & Jaegher, 2015; De Jaegher & Di Paolo, 2007; Di Paolo, Cuffari, & Jaegher, 2018). Although space limitations make any exhaustive comparison impossible for the current paper, we thought it useful to make some tentative comments. This discussion is in response to questions raised, including by an anonymous reviewer, and can be skipped by those who do not wish to pursue a brief discussion of/comparison with enactivism.

Let us first consider the fundamental metaphysics of both frameworks. Autopoietic enactivism (as it has been developed by Di Paolo and colleagues) attempts to derive normativity from the autonomy of the system, which is defined as its ability to sustain identity under precarious conditions. Identity, in turn, is defined as possessing the property of operational closure, which is a property of its organization. By defining autonomy as maintaining identity under "precarious conditions" – recognizing that FFE conditions are necessary for it – Di Paolo and colleagues manage to get something roughly akin to the interactivist property of recursive self-maintenance. However, the proximate source of normativity remains framed in earlier autopoietic terms: It is in reference to the operational closure (a property of system organization) itself that something is considered good or bad for the organism, and not

[21] Interestingly, the special status of social constitution of a person confounds the distinction between epistemology and ontology: In the process of knowing another person, one ipso facto comes to co-constitute the convention that the other person instantiates, as an anticipatory structure that correctly anticipates conventional behavior is also constitutive of that convention. Incidentally, this seems to explain the mutually transformative consequence of cross-cultural interaction – by understanding one another, people *necessarily* become similar. Further analysis is naturally possible: for instance, there seems to be an important difference between "armchair" theoretical knowledge of another culture and embodied knowledge in action of the relevant conventions. These, and more can be explored within the present framework.

[22] Considerable data suggest that nouns are indeed acquired earlier than verbs, even though the acquisition of verbs is itself more nuanced cross-linguistically (see Waxman et al., 2013). However, the present model is not committed to this sequence, depending on the specificity of a given linguistic context, children might follow different pathways even within the same language: for instance, if verb-use dominated around the child, she might start off with verbs and then augment verb use with the addition of nouns.

to its FFE conditions of viability. That is, although Di Paolo (2005) recognizes that FFE conditions *are necessary* for operational closure, he retains the autopoietic idea that self-maintenance is in reference to operational closure and not in relation to such FFE conditions. However, it is not the system's organization per se that is the source of normativity, but rather its relationship to its FFE conditions, and it is this relationship that is being maintained. A FFE system survives if it stays in FFE conditions, not if it exhibits any particular organizational property. Operational closure might indeed be a property of some organizations that self-maintain their relationship to FFE conditions and thus persist, but a given organization is functional or dysfunctional in relationship to such existential conditions, and not to itself or some other organization (see Bickhard, 2016a).

This might seem like a small technical issue, but it does bear onto subsequent development of the enactivist framework. Since it is the system's organization that is being maintained and not its relationship to FFE conditions, change – including enculturation – must either be modeled as reorganization in some inessential "safe region" of the system's organization, whose breakdowns do not obliterate the system's fundamental autonomy (i.e. some essential organizational relationships) (Di Paolo, 2005; cf. Barandiaran & Egbert, 2014) or/and as a proliferation of separate autonomies that interact with each other in ways that allow them to co-exist. Di Paolo et al. (2018) propose the existence of three basic kinds of such separate autonomies – organismic, sensorimotor, and interactional. The dynamics that are claimed to occur between these autonomies form the bedrock of enactivist theorizing, but given the more fundamental issues with the very construct of autonomy we pointed to above, this approach incurs controversy.

The last one of the basic autonomies – interactional – forms the basis for the enactivist model of social cognition and language. An interpersonal encounter is held to "take a life of its own", to generate an independent autonomy, with which the other two autonomies of interacting agents have to grapple. This roughly corresponds to the interactivist emergence of situation convention, but the enactivist account goes into much less detail about how it is achieved – the emergence of such an interpersonal "coupling", as the enactivists term it, is simply assumed to naturally occur in the process of self-organization of the encounter, out of the dialectical tension of the extant autonomies (De Jaegher & Di Paolo, 2007; Di Paolo et al., 2018, pp. 61–86). Since the autonomy of the encounter is regulated by the other autonomies of both agents involved (by their "safe regions"), the agents are claimed to gradually, through a dialectical process, develop ways of co-regulating the autonomy of an encounter in various ways without compromising their own autonomies, a sophisticated form of which amounts to the emergence of language (and enculturation of those agents more generally).

The enactivist idea that language is a means of regulating or interacting with the autonomy of the social encounter is similar to the interactivist proposal that language is a system for interaction with situation convention. And the dynamics and emergent mental organizations that the enactivists claim to be involved there do share the general thrust with what the interactivist model advances. We still see it as questionable, however, that the enactivist proposals builds so centrally on the concept of autonomy. By putting the clash between autonomies of interacting agents in the center of the model (strictly, between their autonomies and the autonomy of their encounter), it might seem that meeting another person needs to pose a grave danger to one's life (organismic autonomy) or freedom (sensorimotor autonomy) in order to incite any attempt at coordination. This conclusion would not be entirely accurate, however, as it is clear that, as it has been developed in the most recent work, partial breakdowns and recoveries of "autonomy" is beginning to do a job similar to interactivist microgenesis, which renders the term itself somewhat misleading – it is not the autonomy (i.e. its ability to self-maintain) of the individual that is being disturbed in a social encounter, but the anticipatory stability of their *representational* (in the interactivist sense, naturally) processes. This function of the notion of autonomy has been embraced when adaptivity was introduced into the enactivist model (Di Paolo, 2005): An autonomous system involves a "safe region", where partial collapses of autonomy are held to allow the system to adapt its functioning to the present situation (cf. Barandiaran & Egbert, 2014). And it is this process that is also recognized in the context of social cognition, for example when Di Paolo et al., 2018, state that "[i]nteractive dissonance can arise in subtle ways and need not be experienced as a major frustration of individual intentions but instead as a discomfort or difficulty in the flow of sensorimotor engagements" (p. 143). This is, we believe, the crux of the matter – social cognition emerges within anticipatory dynamics of interacting agents; it is thanks to anticipatory "agreements" or situation conventions that we understand each other. The notion of autonomy does little to illuminate this process, or even obscures it, as it concerns the internal organization of the system – its operational closure – not its relationship to the social situation. In sum, we see the enactivist notion of autonomy as an unnecessary remnant of autopoietic thinking; it is not needed for normativity – self-maintenance of FFE conditions provides that; and it is not needed to account for social cognition, where it confounds what could otherwise be in our opinion a much clearer discussion of inter-agent anticipatory dynamics.

Another issue is that the term of autonomy seems at times rather equivocated. The authors sometimes discuss autonomy also in a sense that seems much closer to the everyday meaning of the term and seem to posit that interpersonal understanding requires that the interactants' personal freedom be maintained (something that critics of enactivism have picked up on, e.g. Westra & Schönherr, 2019). This is, in our opinion, an unnecessary and mistaken claim that seems to be implicitly invited by the retained autopoietic notion. It is certainly worthwhile to discuss how social conventions constrain individual freedom (e.g. we cannot help but think with the use of our conventional structures), or how lack of freedom of one party influences the emergence of conventions (conventions cannot emerge or would emerge differently if agents are not allowed to freely interact within a certain realm), but these issues are secondary and not fundamental to social cognition or language: A victim of a kidnapping can still understand what his oppressors say as long as he and his oppressors all instantiate the relevant anticipatory conventions, the question of personal freedom is only tangential to the issue of social *cognition*. In short, the notion of autonomy that roughly relates to maintaining one's FFE conditions and the notion of autonomy as high-level personal freedom should not be collapsed: The former indeed is fundamental to social understanding inasmuch as it captures the basic anticipatory nature of cognition at large, but the latter, though related, is a rather different animal. More generally, the charitable application of the notion of autonomy that enactivism gives to an array of different processes, such as metabolism, sensorimotor anticipation, and interactive encounter, carries the danger of equating what could perhaps be best analyzed as different kinds of FFE phenomena.

Certainly, interactivism and enactivism agree on more than they disagree on, and for many purposes the disagreements can be largely ignored – e.g. both frameworks promote research into the dynamics of social encounters as they both hold that it is in such encounters that human sociality emerges in the first place. However, where there are disagreements certainly calls for a more thorough treatment. For this text, we only intend these few remarks to help the reader's appreciation of our proposal if he or she is more familiar with the enactivist framework than interactivism.

5. Current issues in ToM-dominated social cognition research from an interactivist perspective

From the above presentation, it can be easily discerned that interactivism in particular and action-based approaches in general invite different kinds of questions and search for different kinds of answers than the ones of most concern in ToM literature. This can be clearly seen in the existing work of action-based researchers pursuing empirical questions: Issues explored include children's improving representation

of event or social situation schemata (e.g. Dunn, 1988; Nelson, 1986), the emergence of their linguistic abilities within interpersonal and non-linguistic interaction (e.g. Cameron-Faulkner et al., 2020; Carpendale & Carpendale, 2010), or their growing ability to organize their implicit knowledge with the use of reflection (e.g. Nelson & Bruner, 2006). When this research tradition met with the ToM approach, it was initially seen as baffling why ToM researchers are so obsessed with belief attribution and false-belief tasks (see, e.g., Bruner, 1995; Nelson, Plesa, & Henseler, 1998). As should be clear by know, these methodological directions are a natural consequence of the ToM theoretical framework – once you assume that it is all about mental-state attribution, it is the postulated mechanism of attribution you are studying and everything else becomes issues of "mere" performance to be controlled for. That ToM models misguide empirical research is certainly another argument against the framework that further undermines its tenability (Allen & Bickhard, 2013; Ilgaz & Allen, 2020). However, it remains a fact that ToM-(mis)guided research has dominated the field of psychological research on social cognition as well as many philosophical debates of the problem. For that reason, we have decided to offer an explicit treatment of some of the issues of social cognition most widely discussed in ToM literature, with the hope to demonstrate that ToM-based explanations of these phenomena are not the only option available to a researcher, and that they are well accounted for within our preferred framework.

5.1. False-belief task

Given its centrality in contemporary debates on social cognition, representation of beliefs and desires needs to be addressed by our proposal. Within interactivism, propositional attitudes are abstract properties of socio-cognitive structures – they are not elements for constructing such structures. It is indeed useful to abstract the property of predication from a class of conventional structures and differentiate them *reflectively* on the basis of kinds of predications involved – such as believing truly or falsely, or desiring – and children eventually do it. However, this ability is in the current model necessarily derivative from the structures that instantiate such abstract properties and needs to be differentiated from them analytically. The child first forms her level-1 anticipatory representations of social situations, and only then can she start reflectively constructing such level-2 representations as of believing or desiring.[23] Children seem generally capable of such reflective abstraction around age 4 (across different domains, not just in social cognition, see Allen and Bickhard (2018), Human Development Special Issue (1992)).[24]

Let us briefly center on false-belief understanding. Following the spontaneous-response FBT study by Onishi and Baillargeon (2005), many have argued that by 15-months, infants are already able to anticipate false-belief congruent behavior. Recently, however, the robustness of the finding has been questioned as scientists find it difficult to replicate the results (Crivello & Poulin-Dubois, 2017; Dörrenberg, Rakoczy, & Liszkowski, 2018; Kammermeier & Paulus, 2017; Kulke, Duhn, von Schneider, & Rakoczy, 2018; Kulke, Reiβ, Krist, & Rakoczy, 2017; Poulin-Dubois et al., 2018). This is to be expected within our proposal: Some infants might have some coarse-grained situation convention representations that, rather than attributing a false belief, anticipate the perceived agent to behave in accordance with the conventional scheme, which happens to manifest false-belief *congruent* anticipation, but does not involve genuine false-belief understanding. So much is the general interpretation of the spontaneous-response FBT also endorsed by other action-based theorists (e.g. Carpendale & Lewis, 2015); the procedural knowledge of the situation that the infant already possesses anticipates others to *act* in accordance with their false beliefs, but this anticipation is implicit in the infant's embodied or sensori-motor representation of the situation, not a result of the process of belief attribution. The specific form of such possible schemes has been explored by the anti-mentalistic criticism that followed the studies and remains an empirical issue (Borg, 2018; Fenici & Zawidzki, 2016; Heyes, 2014). What we would like to suggest additionally is that at such an early age the conventional part of the child's functional organization is likely to be highly idiosyncratic and contingent on the particularities of the experience in their short life, which could explain the systematic unreliability of the findings.

The elicited-response FBT seems to require *some* reflective abstraction (cf. Kloo, Kristen-Antonow, & Sodian, 2019). The reason for that is that the child is asked to predict verbally what will happen, not to anticipate it in action. Moreover, the question posed to the child is novel enough for them not to have sufficient level-1 organization to answer it correctly without reflection (Allen & Bickhard, 2018).[25] It is important to note, however, that whatever level-2 functional structure the child deploys in the task is likely far from the generality that characterizes an adult folk psychologist; the reflective folk psychological framework is constructed according to the same principles as level-1 functional organization, and can always be potentially extended and revised, including new properties and aspects of the overall functional organization of conventional world knowledge (cf. Carpendale & Lewis, 2006; Nelson, 2007; Stone, Carpendale, Sugarman, & Martin, 2012).

5.2. Folk psychology

Folk psychology within the present framework is a complex phenomenon. Most fundamentally, folk psychology can be defined as a reflective convention. It is reflective in that it involves abstracting certain properties of the functional organization of the mind, such as believing or desiring. And it is a convention in that it is shared by individuals in a society and arguably originated as a solution to various coordination challenges. Overall, anytime commonality in *thinking* (i.e. level-2 anticipation) about people is indispensable for social coordination, folk psychology is the convention we enact. Naturally, there are many ways in which this commonality is achieved, and what falls under the label of folk psychology is likely varied and cognitively differentiated structures that are deployed in various contexts, and towards various goals. Recent work within the so-called pluralist approaches to folk psychology has explored the landscape of these structures (Andrews, 2012, 2015; Fiebich, 2019; Fiebich & Coltheart, 2015; Newen, 2015; Spaulding, 2018). The pluralist approaches to social cognition can be viewed as a culminating point in the long and variegated critique of the ToM conception to social cognition that puts belief-attribution at its base. Drawing on individual points made in earlier criticisms (e.g. Bruner, 1990, 1995; De Jaegher & Di Paolo, 2007; Gallagher, 2001, 2008; Goldman, 2008; Hutto, 2008), these scholars point out that folk psychological reflection does not only involve theorizing about mental states of others, but also stereotyping, explanation by situation, social norms, trait attribution, and potentially more. And these serve a varied set of purposes, including but hardly limited to behavior prediction. This claim is further supported by anthropological analyses that show that folk psychologies differ greatly across cultures

[23] A similar proposal has been offered as a synthesis of ToM and embodiment accounts of social cognition by Michael, Christensen, & Overgaard, 2014.

[24] This is due either to the sufficient stability of structures at level-1 or to some biological maturation of brain architecture that allows for internal reflection around that age. The issue of particular timing is primarily empirical, but the sequence is metaphysically necessary: Level 1 structures have to be constructed first for level 2 to have anything to interact with.

[25] In principle, any anticipation can be successfully accomplished by level 1, but it will always require practice; hence, good performance in interactively novel tasks is a good indicator of reflective processes being involved.

(Lillard, 1998; Luhrmann, 2011; Wierzbicka, 2006). Interactivism provides a framework for analyzing these strategies as different functional structures that participate in the constitution of social reality and social cognition.[26]

Consistent with the pluralist claims, our proposal then accommodates both the traditional view of folk psychology as an explanatory and predictive tool, as well as the more recent points about its different functions. As we have pointed out in our critique in section 2, theory of mind (be it in the form of an implicit mechanism or an actual folk psychological reflection) is generally a bad theory in the sense that it far from exhausts the determinants of another's behavior (and even worse when it is understood as referring to 'things' that don't exist — e.g., encoded propositions in a belief box). However, understood as a reflection over or internal interaction with already contentful socio-cognitive anticipations of social situations, it can aid prediction and explanation. Assume that someone tells me that the man standing in front of my door is the service technician I spoke with over the phone about fixing my broken fridge, and I reflectively attribute to him the mental state of waiting for me and the desire to enter my house and do his job. The folk psychological mental-state attribution modulates my anticipatory processes such that they presuppose him to perhaps greet me and state his business. This presupposition can be wrong or right and is subject to falsification when interacting with the man (e.g. If he is not really the repairman and I start talking to him about my broken fridge and invite him in, his behavior will be unlikely to follow my anticipations. He might try to tell me that I have mistaken him for someone else, or perhaps walk away, giving me a confused look). The important point is that this attribution alone does not explain the whole of my anticipatory processes about that person – the attribution merely modulates the process, but most of implicit presuppositions present there come from general conventionalization of my mind acquired in my culture (e.g. that he will want to shake hands with me or that he has tools in his bag are part of implicit anticipatory structures onto which my explicit attribution only has additional influence, they do not originate in the attribution).

As mentioned earlier, however, behavior prediction is rarely how we use folk psychology and our proposal makes space for such uses as well – folk psychology is a versatile reflective convention that apart from explaining and predicting can be used to communicate one's thoughts, regulate other's behavior, excuse it, and others. One notable point has been made by Brandom (e.g. 2000) and, in the context of ToM research, by Fenici and Zawidzki (2020). The main claim that they champion is that folk psychology is a very bad theory and its real usefulness lies rather in how it regulates social commitments and obligations. For instance, if I say that "John believes X", I am not just or only or at all stating a theoretical claim — I am committing myself to back up that 'claim' if called upon, and/or to justify it (and thus my status as a reliable informative interlocutor) if it turns out to be not correct, and leads you astray, etc. An encoding framework leads one to ignore such social interaction issues, and to interpret folk psychology as a solely denotational theory. Our proposal, on the other hand, can make sense of such uses as well. What the convention of folk psychology does is modulate social anticipations in ways that are shared by people in a given culture (it is a convention); this modulation can be used theoretically to correctly anticipate behavior (function of prediction), but it can also be used regulatively (and perhaps this is its most significant role) to organize social interaction in ways that the above researchers stress. As a reflective convention, folk psychology improves social cognition and coordination, but is not the driving force behind it.

5.3. Social cognition in different cultures

We have pointed out that enculturation of social cognition poses an insurmountable challenge for ToM frameworks. The interactivist model, on the other hand, accommodates enculturation naturally as part of the process of conventionalization of the mind, and provides a framework for studying cultural effects on social cognition.

The most general point is that apperception of a person is largely constituted by the convention of the given culture, which in turn depends on the kinds of coordination challenges the society has faced in the past and the precedence of solutions to those challenges. Naturally, there are culturally universal coordination challenges as all communities interact with physical reality in order to ensure survival. Taking into account the commonality of embeddedness, embodiment, and evolutionary heritage of humans, it is expected that conventions that emerge as solutions to such basic survival challenges will be at least generally comprehensible cross-culturally (e.g. pointing to draw attention, or going hunting).

It is important to understand such cross-cultural universals correctly. As we have already mentioned, they are not some universal, more primitive conceptual foundation upon or out of which culturally specific structures are built, as core knowledge accounts have held (Spelke & Kinzler, 2007). Rather, they are descriptions of stability across different populations and individuals, stability that in the present framework is explained as the outcome of innate scaffolds, and commonalities in embodiment and embeddedness of the constructive process (Carpendale & Wereha, 2013). In short, descriptions of cultural universals should not be reified into their explanation.

When it comes to coordination challenges that are posed in interaction already *within* the realm of social convention – e.g. when a decision has to be made where to go on a hunting trip – then greater arbitrariness is to be expected. For instance, some communities may develop a social hierarchy where the tribe's chief makes the decision; some others might develop a voting convention. Correspondingly, the higher we go in the emergent hierarchy of conventional reality, the bigger cross-cultural variance is likely, which is perhaps best observed in the diversity of languages, ideologies, and spiritual systems of the world's oldest civilizations.

As far as folk psychology is concerned, beliefs and desires will be present implicitly in conventions of all cultures because they are properties of human cognition in general, not of a particular conventional organization – they denote the intrinsic constraints on knowing at large. So, it makes sense that most cultures develop folk psychologies that capture those properties (although arguably not in entirely the same ways). Indeed, all folk psychological concepts identified by Wierzbicka as truly culturally universal designate characteristics of functional organization that are of natural rather than conventional provenance (Wierzbicka, 2005, p. 265; cf.; Wierzbicka, 2006).

There will be, however, a plethora of other organizational properties of social reality and therefore of conventional minds that are increasingly culturally specific, and these too will often figure in a given culture's folk psychology. These culturally specific folk psychological concepts usually capture characteristics of *conventional* minds who exhibit organizational properties that do not exist outside the structure of that convention, or exist but in a highly dissimilar form (e.g. Lomas, 2018; Wierzbicka, 1999). Consider, for instance, such propositional attitudes as *allow, refuse, deny,* or *be disappointed in* – they all presuppose a conventional realm of interaction such that makes it possible for someone to allow someone else to do something, or to be disappointed in someone. Any cognitive system implicitly believes things, but allowing something to happen requires one to be an encultured agent. Even though the general categories of this sort seem relatively common cross-culturally, the specifics of their meaning will vary as the social reality that makes it possible for someone to allow, refuse or deny things

[26] Many of the pluralist accounts remain largely metaphysically agnostic and thus are potentially consistent with the present proposal (Andrews, 2012; Fiebich, 2019; Newen, 2015; Zawidzki, 2013), which is advanced as a "basic-level" metaphysical framework for understanding enculturation. Others make more concrete metaphysical claims and likely clash with the present proposal (Gallagher, 2015).

often differs cross-culturally.

Further, varying historical contingencies and unique ecological conditions might lead societies to develop highly singular conventions. For instance, the already mentioned Welsh concept *hiraeth* signifies longing for an idealized Wales from the past, one that perhaps never existed, but which permeates Welsh mentality (Polk, 1982). Arguably, Welsh minds often engage in activity that instantiates the property of *hiraeth*, and so it has been useful for the culture to abstract it and include it in its folk psychology. Similar, uniquely conventional ontology can be argued for concepts such as the Chinese *xin* 心, which is characteristic of the traditional way of relating to one another in China (Gut, Hryniewska, Pejda, Mirski, & Stoch, 2019; Yu, 2009). These concepts capture the relations that exist within the organization of the social reality specific to their cultures and can be only approximated by an individual who does not instantiate the convention in question.

5.3.1. Functional scaffolding

The concept of *functional scaffolding* allows us to understand more deeply the way in which the child's construction of culturally adequate functional organization is structured by her social environment (Bickhard, 1992a; 1992b, 2005). As has been noted, interactivism models development as recursively constructive; functional organizations that were successful in past interactions are retained in the world knowledge and form the context of further development. This has important consequences: First, certain conventional interactions might require complex constructions that are unlikely to be successfully generated by the agent from scratch. These constructions, however, might be scaffolded by first engaging in interactions that the agent *can* be successful in, which will be retained and will form the context for interaction with the more difficult task, potentially increasing chances of success and formation of the corresponding construction. In other words, the agent might first have to gain some competence or knowledge in order for further learning to have the desired effect (e.g. teaching a toddler how to write is no use when they cannot hold a pen properly, so successfully holding a pen is a good way to start). What happens to the scaffolds once the target construction is achieved might vary – e.g. they might be no longer needed and thus deteriorate, or they might form an intrinsic part of the target construction, effectively constituting the intension of the representation. Clearly, the model provides numerous ways in which this constructive process can be analyzed and studied.

What is important for the present discussion is how the environment introduces such functional scaffolds. Functional scaffolding consists in *blocking of selection pressures* of certain tasks so that children can acquire the right constructions. Such blocking can be achieved in multiple ways, many of which are clearly present in societies across the world. In essence, however, it boils down to imposing certain interaction onto the child, one that the child can be successful in, so that she will form the right constructions that will allow her to achieve success in some other, more complex interaction at a later date. Note that this contrasts with the Vygotskyan notion of scaffolding as provision of knowledge; on an encodingist reading of Vygotsky's ideas, such provision of knowledge must amount to encoding of new information provided by the parent, which we pointed out is impossible in principle. Blocking selection pressures can but *does not have to* be guided by the parent's knowledge of the target construction: What does the scaffolding job proximally is the interactive context that allows for a successful construction, not some transmission of the target structure from the parent's mind. While blocking of selection pressures is often targeted at the child constructing the same knowledge as the parent has, it might just as well be guided by the parent wanting the child to construct knowledge that they *do not* themselves have – they only need to engage the child in the scaffolding interaction. This is nicely illustrated by the work of Reading Corps, an organization that helps to improve children's literacy, which holds trainings for parents who are often illiterate themselves and teaches them to engage in activities that help the child become a better reader (how to hold a book, turn pages, identify letters etc.) (Minnesota, 2020). Generally, the point we wish to make here is that social scaffolding, even though often guided by the scaffolder's knowledge of the target construction, is achieved not by transmission of that knowledge, but by engaging the child in an interaction with a blocked selection pressure(s) that allows her to achieve interactive success and retain the appropriate constructs to be used in further learning and development.[27]

The most prominent example of functional scaffolding is perhaps the formal schooling system – children are not expected to appreciate the aim of education, and would probably not engage in the necessary activities were they not skillfully manipulated by the society to do so. Such scaffolding is necessary since the social reality of modern civilization is constituted by conventions accumulated over millennia, and children would not be able to become participants of it otherwise.

The present framework provides tools for understanding how scaffolding influences the construction and constitution of socio-cognitive structures, or conventional minds. Writing and reading are obvious examples, but there are also less conspicuous cases. Consider, for instance, hand-holding with a parent. In the present framework, it will be a convention that scaffolds a number of other tasks and gets discarded later in life (although it can be co-opted later as a form of interaction in romantic relationships, childrearing as a parent, or stress-coping strategy (Weekes, Kagan, James, & Seboni, 1993)). Turn-taking, on the other hand, is a convention usually scaffolded for children that seems to be retained as part of many other conventions (Vandell & Wilson, 1987), most significantly language (Levinson, 2016).

As explored by some developmentalists, the earliest conventional structures of the child's mind will emerge in dyadic interactions or engagements between the child and their caregiver. Reddy (2018) cites evidence that very early in their development (two and three months), infants exhibit anticipatory behavior in dyadic social situations. For example, they adjust their body when the caregiver begins to pick them up (Reddy, Markova, & Wallot, 2013), which adjustment ceases if the caregiver delays her part of the interaction (Fantasia, Markova, Fasulo, Costall, & Reddy, 2015). From the perspective of our proposal, such enacted routines are earliest conventions that emerge between the child and their caregiver. Since conventions are dynamic structures – they are anticipations of the flow of a social situation and thus are time sensitive; if an interactant behaves out of the anticipated flow, anticipation within the situation convention can break down (i.e. timing is intrinsic in situation conventions), which can explain the effect of delays found by Fantasia et al. (2015). Once such second-personal conventions are stabilized, the child has available conventional anticipatory structures that can scaffold her engagement with triadic situation conventions. For instance, being familiar with the second-personal convention of bottle-feeding, the infant is more likely to apperceive an agent who is feeding *someone else* – for example, a doll – from a third person perspective, initially using the same anticipatory structure but gradually reconstructing it to fit the third-personal flow of interaction. Naturally, successfully observing someone else being fed and being fed oneself

[27] Scaffolding more generally is achieved by introducing developmental constraints in the developmental system that result in the emergence of some system organization (Oyama, 1985/2000). The introduction of the constraints can happen by human design, as is the case in social scaffolding under discussion here, but also naturally, by variation and selective retention of developmental contingencies in phylogeny (importantly, even in the case of human "design", the original emergence of the design is also necessarily by variation and retention). Kuo's (1967) work on embryonic development of pecking in chicks is a nice illustration here. At some point in a chick embryo's development, its heart begins beating and induces head to move in a roughly pecking manner. Kuo found out that blocking this relation resulted in that the chick could not peck, eat, or even remove itself from the shell. Thus, the heart's influence clearly scaffolds the development of the chick's ability to peck. Yet, it is clear that there is no knowledge of pecking that is being transmitted here – it is the modulation of the developmental process created by the heartbeat that does the job.

involve a great deal of distinct anticipatory presuppositions, but the point is that certain anticipatory organizations established by engagement in second-personal feeding can be co-opted in observing someone else being fed. Importantly, they can be co-opted on the basis of organizational similarity of the anticipatory flow and via random variation and selective retention when the second-personal structure fails to anticipate a third-personal interaction successfully – no homuncular agent doing the comparing within the organism is needed.

Another example, related to reflective abstraction, is reading books for children that simplify social relations, and commenting on them with the child, making sure they understand them well with such techniques as relating the events to the child's own life (Farkas et al., 2020; Salo, Rowe, Leech, & Cabrera, 2016; Ziv, Smadja, & Aram, 2014). In this case, scaffolding seems to concern mostly the child's reflection over social reality, and construction of explicit, folk psychological concepts that can be applied to their own experience. Indeed, such book-reading does predict success on explicit-reasoning ToM tasks (Adrián, Clemente, & Villanueva, 2007; Symons, Peterson, Slaughter, Roche, & Doyle, 2005). A discussion related to this kind of scaffolding follows in the next subsection.

5.4. How does language influence socio-cognitive development?

The convention of language is certainly the most pervasive constituent of social reality; it not only is the most powerful way to interact with situation conventions, but it also constitutes many of them. As such, much of the content of the human mind is language-constituted,[28] and perhaps an even larger part of it is implicated in language in the sense that there are functional links from most of functional structures (e.g. representations of objects) to the convention of language (as long as we can talk about something, there must be such a functional link). Therefore, within the proposed framework, asking how language relates to social cognition is akin to asking how vision relates to cognition of physical reality: We know others to a great extent through (anticipation of) linguistic interaction similarly to how we know the physical world significantly through (anticipation of) visual interaction (cf. Gibson's ecological account of vision).

This perspective contrasts radically with the one of the traditional ToM frameworks. As Ilgaz and Allen (2020) have recently pointed out, traditional ToM models necessarily view language in an instrumental way – it is a source of data for the construction of theory of mind and as such can aid the process of its development, but it does not partake in the constitution of social cognition in any way. The authors highlight that this instrumental perspective imposed by the ontological commitments of the traditional frameworks has (mis)guided most of the research on language's role in socio-cognitive development. Linguistic interaction has been looked at primarily as a source of information about the abstract concepts of believing and desiring. This is reflected in both what aspect of linguistic interaction is seen as methodologically relevant (e.g. frequency of use of the terms 'think' or 'want' in the household), as well as what is seen as evidence of improved socio-cognitive skills (mostly false-belief task performance) (cf. Devine & Hughes, 2018; Milligan, Astington, & La Dack, 2007; Tompkins, Benigno, Kiger Lee, & Wright, 2018). From the standpoint of our proposal, although important, this is only a specific part of what social cognition is and how language matters for it.

Naturally, although limited in its scope, the evidence is valuable and requires an interpretation from the present perspective. We will focus specifically on the impact of mastering the terms of "think" and "want", which has been found to improve ToM understanding (Furrow, Moore, Davidge, & Chiasson, 1992; Gola, 2012; Roby & Scott, 2018; Ruffman, Puri, Galloway, Su, & Taumoepeau, 2018; Ruffman, Slade, & Crowe, 2002; Ruffman, Slade, Rowlandson, Rumsey, & Garnham, 2003), but the points are generalizable to the influence of folk psychological talk in general.

As noted earlier, believing and desiring understood broadly is instantiated by living beings in general – implicit presupposition is a form of believing, and goal-orientedness is a form of desiring. We propose that mastering such terms as 'think' or 'want' creates a functional organization that links classes of functional structures that instantiate such abstract relations with possibilities for social interaction (with conventional structures); it makes the child cognitively differentiate functional organization on the basis of belonging to those abstract classes (something which is not interactively relevant outside of social reality and would probably not be differentiated without language, cf. Pyers and Senghas (2009)). This claim can be perhaps made clearer by a comparison with an analogous differentiation but in a more familiar domain. Take the previously used example of the edible fruit that is sweet and soft and can be identified by squeezing it (as opposed to the other, inedible fruit that is firm, bitter, and makes one nauseous, but which is identical to the former in all its other characteristics). Cognitive differentiation between softness and firmness of the fruit makes a difference for the organism – by learning to differentiate the two kinds of fruit on the basis of their property of softness,[29] the organism can better interact with its environment (i.e. it can achieve anticipatory success). Now, our suggestion is that a similar advantage for anticipatory success in a cultural milieu is bestowed by differentiating conventional cognitive structures on the basis of the property of belief and desire. These properties, although inherently present in any real-life cognitive agent (their situation knowledge), can be only differentiated if they make a difference for how the process of interaction flows. They make such a difference most prominently in conversations about mental states where mental-state talk corresponds to how people behave (whether somebody informs us of someone else's belief, or they declare it themselves, or they hold us accountable for a belief we have voiced earlier, or the child's caregiver attributes a belief to a character in the picture book they are reading etc.), and so learning to successfully anticipate the flow of such a mental-state interaction requires – and thus provides an incentive to construct – differentiations of conventional structures on the basis of mental states held by the interactants. This is naturally a much more complex task than the one in the fruit example, but mentalistic language certainly helps it – once mastered, it creates an anticipatory organization in which such a differentiation is embodied. In order to correctly use and understand folk psychological talk, the child needs to establish appropriate functional link on the basis of such a differentiation; otherwise, she will fail in her anticipation. Importantly, this *does not* imply that believing and desiring are represented reflectively or explicitly, but only that they become an organizing principle of the functional organization of the child's mind when she learns to communicate effectively with the use of such terms.

That being said, a full competence in folk psychology seems to require appropriate reflective functional organization as well – being able to apply the principles of folk psychology to novel examples, to think back to past situations, to compare and hypothesize about different possible mental states etc. are all frequent presuppositions of a conversation about social life. Without thinking reflectively about believing and desiring, there seems to be little possibility of anticipatory success in such conversations. Consequently, folk psychological conversations can be seen as imposing a selection pressure onto the reflective organization of the child, and once the child is capable of reflection, engaging in joint contemplation of minds with the parent provides more opportunities for the generation and subsequent

[28] Remember that content is meant here in the interactivist sense: It is constituted as implicit presuppositions of anticipatory processes, so when there is anticipation of linguistic interaction, the content is language-constituted.

[29] Strictly, on the basis of two possible process flows of squeezing that implicitly presuppose either softness or firmness of the object and are part of the organism's functional structures that implicitly represent the fruit.

retention of the right reflective structures (note that the process is usually scaffolded by the parent, as mentioned earlier).

Thus, as an initial claim to be extended in the future, we wish to point to two ways in which mental-state talk likely contributes to the child's improved competence in reflective socio-cognitive tasks: (1) being able to use such terms as 'think' and 'want' for varied interactional purposes results in a functional organization that differentiates classes of structures on the basis of their abstract properties, such as believing and desiring (but also allowing, refusing etc. – this can apply to any abstract property).[30] And (2) being able to engage in folk psychological discussion about other minds (e.g. using mental-state terms in the context of a joint reflective activity) imposes selection pressure onto the child's reflective processes, and thus the more frequent such activity, and the better it is scaffolded by the parent, the quicker the child should construct the appropriate reflective structures.

6. Conclusions

We have argued that ToM models cannot account for enculturation of social cognition in principle due to their encodingist commitments. To make our case, we reviewed three related issues that encodingist frameworks face: the copy problem, foundationalism, and the frame problems. The copy problem and foundationalism imply the impossibility of genuine enculturation: With no independent epistemic access to social reality, a ToM agent has no way to acquire culturally specific concepts, which forces ToM accounts into the claim of innate foundational content that can serve as the basis for development. The frame problems of holism and interpersonal coordination unveil the epistemological inadequacy of mental state attribution as a behavior prediction and coordination ensuring tool: With no model of implicitness, ToM models cannot account for the unbounded relevancy relationships that hold between one's behavior, their mental states, and elements of the world.

As a positive alternative, we have argued for interactivism as a framework for understanding social cognition and its enculturation. We have pointed out its strong metaphysical base (representation as implicit presupposition of anticipatory processes, and normativity rooted in FFE dynamics), and argued that it naturally accommodates enculturation of social cognition, drawing relevant connections with existing empirical and theoretical research. We suggest interactivism has great potential for studying encultured or conventional minds.

Funding information

This work was supported the National Science Centre (Poland) under Grant PRELUDIUM 12 (UMO-2016/23/N/HS1/02887) awarded to Robert Mirski for the research project entitled "On Socio-Cultural Entanglement of Folk Theories of Mind. A Discussion on the Boundaries of Universality and Commonality of Mental Concepts and Attributions."

Declaration of competing interest

The authors declare no potential conflict of interest.

[30] Notice that this nicely accommodates the data showing that non-mentalistic use of "think" and "want" does not help the child do better on ToM tests, and even seem to retard it (S. Lewis, Hacquard, & Lidz, 2012). Mastering such non-mentalistic conversational uses of the terms creates a different functional organization than mastering them in their mentalistic use, and this functional organization, once extant in the mind, might require *re*construction to accommodate mentalistic use of the terms.

References

Adams, R. B., Rule, N. O., Franklin, R. G., Wang, E., Stevenson, M. T., Yoshikawa, S., & Ambady, N. (2010). Cross-cultural reading the mind in the eyes: An fMRI investigation. *Journal of Cognitive Neuroscience, 22*(1), 97–108. https://doi.org/10.1162/jocn.2009.21187

Adrián, J. E., Clemente, R. A., & Villanueva, L. (2007). Mothers' use of cognitive state verbs in picture-book reading and the development of children's understanding of mind: A longitudinal study. *Child Development, 78*(4), 1052–1067. https://doi.org/10.1111/j.1467-8624.2007.01052.x

Allen, J. W. P., & Bickhard, M. H. (2013). Stepping off the pendulum: Why only an action-based approach can transcend the nativist–empiricist debate. *Cognitive Development, 28*(2), 96–133. https://doi.org/10.1016/j.cogdev.2013.01.002

Allen, J. W. P., & Bickhard, M. H. (2018). Stage fright: Internal reflection as a domain general enabling constraint on the emergence of explicit thought. *Cognitive Development, 45*, 77–91. https://doi.org/10.1016/j.cogdev.2017.12.005

Andrews, K. (2012). *Do apes read minds?: Toward a new folk psychology/Kristin Andrews*. Cambridge, Mass., London: MIT Press.

Andrews, K. (2015). The folk psychological spiral: Explanation, regulation, and language. *The Southern Journal of Philosophy, 53*(4), 50–67. https://doi.org/10.1111/sjp.12121

Barandiaran, X. E., & Egbert, M. D. (2014). Norm-establishing and norm-following in autonomous agency. *Artificial Life, 20*(1), 5–28. https://doi.org/10.1162/ARTL_a_00094

Baron-Cohen, S., Tager-Flusberg, H., & Lombardo, M. (2013). Understanding other minds: What's new? In S. Baron-Cohen, H. Tager-Flusberg, & M. Lombardo (Eds.), *Understanding other minds: Perspectives from developmental social neuroscience/edited by Simon baron-cohen, helen tager-flusberg and Michael Lombardo* (3rd ed., pp. v–x).

Berger, P. L., & Luckmann, T. (1967). The social construction of reality: A treatise in the sociology of knowledge. In *Anchor books: A 589* (Anchor books ed.). Garden City, NY: Doubleday.

Bickhard, M. H. (in preparation). The whole person: Toward a naturalism of persons: Contributions to an ontological psychology.

Bickhard, M. H. (1980). *Cognition, convention, and communication*. New York N.Y.: Praeger.

Bickhard, M. H. (1992a). How does the environment affect the person? In L. T. Winegar, & J. Valsiner (Eds.), *Children's development within social context* (pp. 63–92). Hillsdale, N.J.: L. Erlbaum.

Bickhard, M. H. (1992b). Scaffolding and self-scaffolding: Central aspects of development. In L. T. Winegar, & J. Valsiner (Eds.), *Children's development within social context* (pp. 33–52) (Hillsdale, N.J.: L. Erlbaum).

Bickhard, M. H. (1998). Levels of representationality. *Journal of Experimental & Theoretical Artificial Intelligence, 10*(2), 179–215. https://doi.org/10.1080/095281398146824

Bickhard, M. H. (2000). Motivation and emotion: An interactive process model. In R. D. Ellis, & N. Newton (Eds.), *The caldron of consciousness*. Amsterdam: John Benjamins Publishing Company.

Bickhard, M. H. (2001). Why children don't have to solve the frame problems: Cognitive representations are not encodings. *Developmental Review, 21*(2), 224–262. https://doi.org/10.1006/drev.2000.0521

Bickhard, M. H. (2003). Process and emergence: Normative function and representation. In J. Seibt (Ed.), *Process theories* (pp. 121–155). Dordrecht: Springer Netherlands.

Bickhard, M. H. (2005). Functional scaffolding and self-scaffolding. *New Ideas in Psychology, 23*(3), 166–173. https://doi.org/10.1016/j.newideapsych.2006.04.001

Bickhard, M. H. (2007). Language as an interaction system. *New Ideas in Psychology, 25*(2), 171–187. https://doi.org/10.1016/j.newideapsych.2007.02.006

Bickhard, M. H. (2008). Social ontology as convention. *Topoi, 27*(1–2), 139–149. https://doi.org/10.1007/s11245-008-9036-1

Bickhard, M. H. (2009). The interactivist model. *Synthese, 166*(3), 547–591. https://doi.org/10.1007/s11229-008-9375-x

Bickhard, M. H. (2010). Interactive knowing: The metaphysics of intentionality. In R. Poli, & J. Seibt (Eds.), *Theory and applications of ontology: Philosophical perspectives/edited by roberto poli and johanna seibt* (pp. 207–229). Dordrecht, London: Springer.

Bickhard, M. H. (2015a). The social-interactive ontology of language. *Ecological Psychology, 27*(3), 265–277. https://doi.org/10.1080/10407413.2015.1068656

Bickhard, M. H. (2015b). What could cognition be if not computation…Or connectionism, or dynamic systems? *Journal of Theoretical & Philosophical Psychology, 35*(1), 53–66. https://doi.org/10.1037/a0038059

Bickhard, M. H. (2015c). Toward a model of functional brain processes II: Central nervous system functional macro-architecture. *Axiomathes, 25*(4), 377–407. https://doi.org/10.1007/s10516-015-9276-9

Bickhard, M. H. (2015d). Toward a model of functional brain processes I: Central nervous system functional micro-architecture. *Axiomathes, 25*(3), 217–238. https://doi.org/10.1007/s10516-015-9275-x

Bickhard, M. H. (2016a). Inter- and En- activism: Some thoughts and comparisons. *New Ideas in Psychology, 41*, 23–32. https://doi.org/10.1016/j.newideapsych.2015.12.002

Bickhard, M. H. (2016b). The anticipatory brain: Two approaches. In V. C. Müller (Ed.), *Fundamental issues of artificial intelligence* (pp. 259–281). Cham: Springer International Publishing.

Bickhard, M. H., & Campbell, R. L. (1996). Topologies of learning and development. *New Ideas in Psychology, 14*(2), 111–156. https://doi.org/10.1016/0732-118X(96)00015-3

Bickhard, M. H., & Richie, D. M. (1983). *On the nature of representation: A case study of James gibson's theory of perception/mark H. Bickhard, D. Michael Richie with Robert*

Hughes, James dannemiller. New York, NY, USA: Praeger. https://doi.org/10.1057/9781137400437.0005
Bickhard, M. H., & Terveen, L. (1995). *Foundational issues in artificial intelligence and cognitive science: Impasse and solution. Advances in psychology* (Vol. 109). Amsterdam, New York: Elsevier.
Borg, E. (2018). On deflationary accounts of human action understanding. *Review of Philosophy and Psychology, 9*(3), 503–522. https://doi.org/10.1007/s13164-018-0386-3
Brandom, R. (2000). *Making it explicit: Reasoning, representing, and discursive commitment* (3rd pr). Cambridge, Mass: Harvard University Press [u.a.].
Brette, R. (2019). Is coding a relevant metaphor for the brain? *Behavioral and Brain Sciences*, 1–44. https://doi.org/10.1017/S0140525X19000049
Bruner, J. S. (1990). *Acts of meaning. The Jerusalem-Harvard lectures*. Cambridge, Mass: Harvard University Press.
Bruner, J. S. (1995). Commentary. *Human Development, 38*(4–5), 203–213. https://doi.org/10.1159/000278316
Bushnell, I. W. R. (2001). Mother's face recognition in newborn infants: Learning and memory. *Infant and Child Development, 10*(1–2), 67–74. https://doi.org/10.1002/icd.248
Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language, 28*(5), 606–637. https://doi.org/10.1111/mila.12036
Cameron-Faulkner, T., Malik, N., Steele, C., Coretta, S., Serratrice, L., & Lieven, E. (2020). A Cross-Cultural Analysis of Early Prelinguistic Gesture Development and Its Relationship to Language Development. Child Development. Advance online publication. https://doi.org/10.1111/cdev.13406.
Campbell, D. T. (1974). Evolutionary epistemology. In P. A. Schilpp (Ed.), *The philosophy of karl popper* (pp. 413–463). Open Court.
Campbell, R. J. (2015). *The metaphysics of emergence*. New York: Palgrave Macmillan.
Campbell, R. L., & Bickhard, M. H. (1986). *Knowing levels and developmental stages. Contributions to human development* (Vol. 16). Basel: S. Karger.
Carpendale, J. I. M., & Carpendale, A. B. (2010). The development of pointing: From personal directedness to interpersonal direction. *Human Development, 53*(3), 110–126. https://doi.org/10.1159/000315168
Carpendale, J. I. M., & Lewis, C. (2006). *How children develop social understanding. Understanding children's worlds*. Malden, Mass., Oxford: Blackwell Publishers.
Carpendale, J. I. M., & Lewis, C. (2015). The development of children's understanding of social interaction. In S. Robson, & S. Flannery Quinn (Eds.), *Routledge international handbooks of education. The Routledge international handbook of young childrens thinking and understanding* (pp. 123–132). Milton Park, Abingdon, Oxon, New York, NY: Routledge.
Carpendale, J. I. M., & Wereha, T. J. (2013). Understanding common developmental timetables across cultures from a developmental systems perspective. *Human Development, 56*(3), 207–212. https://doi.org/10.1159/000351276
Carruthers, P. (2013). Mindreading in infancy. *Mind & Language, 28*(2), 141–172. https://doi.org/10.1111/mila.12014
Carruthers, P. (2015). *Mindreading in adults: Evaluating two-systems views. Synthese*. Advance online publication. https://doi.org/10.1007/s11229-015-0792-3
Christensen, W. D., & Bickhard, M. H. (2002). The process dynamics of normative function. *The Monist, 85*(1), 3–28.
Christopher, J. C., & Bickhard, M. H. (2007). Culture, self and identity: Interactivist contributions to a metatheory for cultural psychology. *Culture & Psychology, 13*(3), 259–295. https://doi.org/10.1177/1354067X07079881
Colonnesi, C., Stams, G. J. J. M., Koster, I., & Noom, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review, 30*(4), 352–366. https://doi.org/10.1016/j.dr.2010.10.001
Crivello, C., & Poulin-Dubois, D. (2017). *Infants' false belief understanding: A non-replication of the helping task*. Cognitive Development. Advance online publication. https://doi.org/10.1016/j.cogdev.2017.10.003
Cuffari, E. C., Di Paolo, E., & Jaegher, H. de (2015). From participatory sense-making to language: There and back again. *Phenomenology and the Cognitive Sciences, 14*(4), 1089–1125. https://doi.org/10.1007/s11097-014-9404-9
De Bruin, L. C., & Kästner, L. (2012). Dynamic embodied cognition. *Phenomenology and the Cognitive Sciences, 11*(4), 541–563. https://doi.org/10.1007/s11097-011-9223-1
De Jaegher, H., & Di Paolo, E. A. (2007). Participatory sense-making. *Phenomenology and the Cognitive Sciences, 6*(4), 485–507. https://doi.org/10.1007/s11097-007-9076-9
Devine, R. T., & Hughes, C. (2018). Family correlates of false belief understanding in early childhood: A meta-analysis: A meta-analysis. *Child Development, 89*(3), 971–987. https://doi.org/10.1111/cdev.12682
Di Paolo, E. A. (2005). Autopoiesis, Adaptivity, Teleology, Agency. Phenomenology and the. *Cognitive Sciences, 4*(4), 429–452. https://doi.org/10.1007/s11097-005-9002-y.
Di Paolo, E. A., Cuffari, E. C., & Jaegher, H. de (2018). *Linguistic bodies: The continuity between life and language*. Cambridge MA: The MIT Press.
Dixson, H. G. W., Komugabe-Dixson, A. F., Dixson, B. J., & Low, J. (2017). *Scaling theory of mind in a small-scale society: A case study from Vanuatu*. Child Development. Advance online publication. https://doi.org/10.1111/cdev.12919
Doherty, M. J., & Perner, J. (2020). Mental files: Developmental integration of dual naming and theory of mind. *Developmental Review, 56*, 100909. https://doi.org/10.1016/j.dr.2020.100909.
Dörrenberg, S., Rakoczy, H., & Liszkowski, U. (2018). *How (not) to measure infant Theory of Mind: Testing the replicability and validity of four non-verbal measures*. Cognitive Development. Advance online publication. https://doi.org/10.1016/j.cogdev.2018.01.001
Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes/fred dretske*. Cambridge, Mass: MIT Press.
Dunn, J. (1988). *The beginnings of social understanding*. Cambridge, Mass: Harvard University Press.
Eck, D. (2015). *The encultured mind: From cognitive science to social epistemology*. Ph.D. thesis). University of South Florida.
Edelman, S. (2003). But will it scale up? Not without representations. *Adaptive Behavior, 11*(4), 273–275. https://doi.org/10.1177/1059712303114009
Fantasia, V., Markova, G., Fasulo, A., Costall, A., & Reddy, V. (2015). Not just being lifted: Infants are sensitive to delay during a pick-up routine. *Frontiers in Psychology, 6*, 2065. https://doi.org/10.3389/fpsyg.2015.02065
Farkas, C., Santelices, M. P., Vallotton, C. D., Brophy-Herb, H. E., Iglesias, M., Sieverson, C., & Álvarez, C. (2020). Children's storybooks as a source of mental state references: Comparison between books from Chile, Colombia, Scotland and USA. *Cognitive Development, 53*, 100845. https://doi.org/10.1016/j.cogdev.2019.100845
Fenici, M. (2017a). *Rebuilding the landscape of psychological understanding after the mindreading war*. Advance online publication. https://doi.org/10.13128/Phe_Mi-21113
Fenici, M., & Garofoli, D. (2017). The biocultural emergence of mindreading: Integrating cognitive archaeology and human development. *Journal of Cultural Cognitive Science, 1*(2), 89–117. https://doi.org/10.1007/s41809-017-0008-0
Fenici, M., & Zawidzki, T. W. (2016). Action understanding in infancy: Do infant interpreters attribute enduring mental states or track relational properties of transient bouts of behavior? *Studia Philosophica Estonica, 9*(2), 237–257.
Fenici, M., & Zawidzki, T. W. (2020). *The origins of mindreading: How interpretive socio-cognitive practices get off the ground. Synthese*. Advance online publication. https://doi.org/10.1007/s11229-020-02577-4
Fiebich, A. (2019). In defense of pluralist theory. *Synthese, 165*(1), 13. https://doi.org/10.1007/s11229-019-02490-5
Fiebich, A., & Coltheart, M. (2015). Various ways to understand other minds: Towards a pluralistic approach to the explanation of social understanding. *Mind & Language, 30*(3), 235–258. https://doi.org/10.1111/mila.12079
Fodor, J. A. (1975). *The language of thought. The language and thought series*. New York: Crowell.
Fodor, J. A. (1981). The present status of the innateness controversy. In *Representations: Philosophical essays on the foundations of cognitive science/jerry A. Fodor. Brighton*. Harvester.
Fodor, J. A. (1992). A theory of the child's theory of mind. *Cognition, 44*(3), 283–296. https://doi.org/10.1016/0010-0277(92)90004-2
Ford, K. M., & Pylyshyn, Z. W. (1996). The robot's dilemma revisited: The frame problem in artificial intelligence/. In *Kenneth M. Ford & Zenon Pylyshyn. Theoretical issues in cognitive science*. Norwood, NJ: Ablex Pub.
Furrow, D., Moore, C., Davidge, J., & Chiasson, L. (1992). Mental terms in mothers' and children's speech: Similarities and relationships. *Journal of Child Language, 19*, 617. https://doi.org/10.1017/S0305000900011594, 03.
Gallagher, S. (2001). The practice of mind: Theory, simulation or primary interaction? *Journal of Consciousness Studies, 8*(5–7), 83–108.
Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition, 17*(2), 535–543. https://doi.org/10.1016/j.concog.2008.03.003
Gallagher, S. (2015). The new hybrids: Continuing debates on social perception. *Consciousness and Cognition, 36*, 452–465. https://doi.org/10.1016/j.concog.2015.04.002
Geertz, C. (1973). *The interpretation of cultures; selected essays*.
Gola, A. A. H. (2012). Mental verb input for promoting children's theory of mind: A training study. *Cognitive Development, 27*(1), 64–76. https://doi.org/10.1016/j.cogdev.2011.10.003
Goldinger, S. D., Papesh, M. H., Barnhart, A. S., Hansen, W. A., & Hout, M. C. (2016). The poverty of embodied cognition. *Psychonomic Bulletin & Review, 23*(4), 959–978. https://doi.org/10.3758/s13423-015-0860-1
Goldman, A. I. (2008). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading. Philosophy of mind*. Oxford, New York: Oxford University Press.
Gopnik, A. (2003). The theory theory as an alternative to the innateness hypothesis. In L. M. Antony, & N. Hornstein (Eds.), *Philosophers and their critics. Chomsky and his critics* (pp. 238–254). Malden, MA: Blackwell Pub. https://doi.org/10.1002/9780470690024.ch10.
Gopnik, A. (2011). The theory theory 2.0: Probabilistic models and cognitive development. *Child Development Perspectives, 5*(3), 161–163. https://doi.org/10.1111/j.1750-8606.2011.00179.x
Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language, 7*(1–2), 145–171. https://doi.org/10.1111/j.1468-0017.1992.tb00202.x
Groome, D., & Brace, N. (2014). *An introduction to cognitive psychology: Processes and disorders* (3rd ed.). London, New York: Psychology Press, Taylor & Francis Group.
Gut, A., Hryniewska, J., Pejda, K., Mirski, R., & Stoch, N. (2019). Body and mind: Comparative research on mentalistic conceptualization in Poland and China. In B. Duda, R. Kiełtyka, & E. Konieczna (Eds.), *Studies in linguistic, anglophone literatures and cultures* (pp. 70–92). Peter Lang.
Helming, K. A., Strickland, B., & Jacob, P. (2014). Making sense of early false-belief understanding. *Trends in Cognitive Sciences, 18*(4), 167–170. https://doi.org/10.1016/j.tics.2014.01.005
Helming, K. A., Strickland, B., & Jacob, P. (2016). Solving the puzzle about early belief-ascription. *Mind & Language, 31*(4), 438–469. https://doi.org/10.1111/mila.12114
Heyes, C. (2014). Submentalizing: I Am not really reading your mind. *Perspectives on psychological science. A Journal of the Association for Psychological Science, 9*(2), 131–143. https://doi.org/10.1177/1745691613518076
Heyes, C. (2018). *Cognitive gadgets: The cultural evolution of thinking*. Cambridge Massachusetts: The Belknap Press of Harvard University Press.
Hughes, C., Devine, R. T., Ensor, R., Koyasu, M., Mizokawa, A., & Lecce, S. (2014). Lost in translation?: Comparing British, Japanese, and Italian children's theory-of-mind

performance. *Child Development Research, 2014*(6), 1–10. https://doi.org/10.1155/2014/893492
Human Development Special Issue. (1992). The age 4 transition [special issue]. *Human Development, 35*(3).
Hutto, D. D. (2008). *Folk psychological narratives: The sociocultural basis of understanding reasons/daniel D. Hutto*. Cambridge, Mass., London: MIT Press.
Hutto, D. D., & Myin, E. (2013). *Radicalizing enactivism: Basic minds without content/daniel D. Hutto and erik Myin*. Cambridge, MA: MIT Press.
Hutto, D. D., & Myin, E. (2017). *Evolving enactivism: Basic minds meet content*. Cambridge Massachusetts: MIT Press.
Ilgaz, H., & Allen, J. W. P. (2020). (Co-)Constructing a theory of mind: From language or through language? *Synthese, 78*(3), 1–22. https://doi.org/10.1007/s11229-020-02581-8
Kammermeier, M., & Paulus, M. (2017). Do action-based tasks evidence false-belief understanding in young children? *Cognitive development. Advance online publication, 46*, 31–39. https://doi.org/10.1016/j.cogdev.2017.11.004
Kempson, R., Cann, R., Gregoromichelaki, E., & Chatzikyriakidis, S. (2016). Language as mechanisms for interaction. *Theoretical Linguistics, 42*(3–4), 1. https://doi.org/10.1515/tl-2016-0011
Kempson, R., Cann, R., Gregoromichelaki, E., & Chatzikyriakidis, S. (2017). Action-based grammar. *Theoretical Linguistics, 43*(1–2), 643672. https://doi.org/10.1515/tl-2017-0012
Kloo, D., Kristen-Antonow, S., & Sodian, B. (2019). Progressing from an implicit to an explicit false belief understanding: A matter of executive control? *International Journal of Behavioral Development, 85*. https://doi.org/10.1177/0165025419850901, 016502541985090.
Kobayashi, C., Glover, G. H., & Temple, E. (2006). Cultural and linguistic influence on neural bases of 'Theory of Mind': An fMRI study with Japanese bilinguals. *Brain and Language, 98*(2), 210–220. https://doi.org/10.1016/j.bandl.2006.04.013
Kulke, L., Duhn, B., von Schneider, D., & Rakoczy, H. (2018). Is implicit theory of mind a real and robust phenomenon? Results from a systematic replication study. *Psychological Science, 29*(6), 888–900. https://doi.org/10.1177/0956797617747090
Kulke, L., Reiß, M., Krist, H., & Rakoczy, H. (2017). *How robust are anticipatory looking measures of theory of mind? Replication attempts across the life span. Cognitive Development.* Advance online publication. https://doi.org/10.1016/j.cogdev.2017.09.001
Kuo, Z.-Y. (1967). *The dynamics of behavior development: An epigenetic view*. New York: Random House.
Lavelle, J. S. (2019). The impact of culture on mindreading. *Synthese, 87*(3), 1–24. https://doi.org/10.1007/s11229-019-02466-5
Legros, S., & Cislaghi, B. (2020). Mapping the social-norms literature: An overview of reviews. *Perspectives on psychological science. A Journal of the Association for Psychological Science, 15*(1), 62–80. https://doi.org/10.1177/1745691619866455
Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in "theory of mind". *Trends in Cognitive Sciences, 8*(12), 528–533. https://doi.org/10.1016/j.tics.2004.10.001
Levinson, S. C. (2016). Turn-taking in human communication—origins and implications for language processing. *Trends in Cognitive Sciences, 20*(1), 6–14. https://doi.org/10.1016/j.tics.2015.10.010
Lewis, D. K. (2002). *Convention: A philosophical study* [1969]). Oxford: Blackwell.
Lewis, S., Hacquard, V., & Lidz, J. (2012). The semantics and pragmatics of belief reports in preschoolers. *Semantics and Linguistic Theory, 22*, 247. https://doi.org/10.3765/salt.v22i0.3085
Lillard, A. S. (1998). Ethnopsychologies: Cultural variations in theories of mind. *Psychological Bulletin, 123*(1), 3–32. https://doi.org/10.1037/0033-2909.123.1.3
Lomas, T. (2018). *Translating happiness: A cross-cultural lexicon of well-being*. Cambridge, Massachusetts: The MIT Press.
Luhrmann, T. (2011). Towards an anthropological theory of mind. *Journal of the Finnish Anthropological Society, 36*, 5–69.
Mayer, A., & Träuble, B. (2012). Synchrony in the onset of mental state understanding across cultures?: A study among children in Samoa. *International Journal of Behavioral Development, 37*(1), 21–28. https://doi.org/10.1177/0165025412454030
McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science (New York, N.Y.), 317*(5838), 631. https://doi.org/10.1126/science.1144073
Meltzoff, A. N., & Gopnik, A. (2013). Learning about the mind from evidence: Children's development of intuitive theories of perception and personality. In S. Baron-Cohen, H. Tager-Flusberg, & M. Lombardo (Eds.), *Understanding other minds: Perspectives from developmental social neuroscience/edited by Simon baron-cohen, helen tager-flusberg and Michael Lombardo* (3rd ed., pp. 19–34). https://doi.org/10.1093/acprof:oso/9780199692972.003.0002
Michael, J., Christensen, W., & Overgaard, S. (2014). Mindreading as social expertise. *Synthese, 191*(5), 817–840. https://doi.org/10.1007/s11229-013-0295-z
Milligan, K., Astington, J. W., & La Dack. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development, 78*(2), 622–646. https://doi.org/10.1111/j.1467-8624.2007.01018.x
Millikan, R. G. (1984). *Language, thought, and other biological categories: New foundations for realism*. Cambridge, Mass: MIT Press. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=49593.
Millikan, R. G. (1995). *White queen psychology and other essays for Alice (2. print). A Bradford book*. Cambridge, Mass: MIT Press.
Minnesota reading Corps website. Retrieved from https://minnesotareadingcorps.org/, (2020, March 23).
Mirski, R., & Bickhard, M. H. (2019). Encodingism is not just a bad metaphor. *Behavioral and Brain Sciences, 42*, e237. https://doi.org/10.1017/S0140525X19001286
Mirski, R., Bickhard, M. H., Eck, D., & Gut, A. (2020). Encultured minds, not error reduction minds. *Behavioral and Brain Sciences, 43*, e109. https://doi.org/10.1017/S0140525X19002826
Mirski, R., & Gut, A. (2018). Action-based versus cognitivist perspectives on socio-cognitive development: Culture, language and social experience within the two paradigms. *Synthese, 28*(2), 96. https://doi.org/10.1007/s11229-018-01976-y
Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron, 50*(4), 655–663. https://doi.org/10.1016/j.neuron.2006.03.040
Morton, A. (1996). Folk psychology is not a predictive device. *Mind, 105*(417), 119–137. https://doi.org/10.1093/mind/105.417.119
Nelson, K. (1986). *Event knowledge: Structure and function in development*. Hillsdale, NJ: Erlbaum.
Nelson, K. (2007). *Young minds in social worlds: Experience, meaning, and memory/Katherine Nelson*. Cambridge, Mass., London: Harvard University Press.
Nelson, K., & Bruner, J. S. (2006). Narratives from the crib. In *University Press paperback ed* (1st Harvard). Cambridge, Mass., London: Harvard University Press.
Nelson, K., Plesa, D., & Henseler, S. (1998). Children's theory of mind: An experiential interpretation. *Human Development, 41*(1), 7–29. https://doi.org/10.1159/000022565
Newell, A., & Simon, H. A. (2006). Computer science as empirical enquiry: Symbols and search. In J. L. Bermúdez (Ed.), *Routledge contemporary readings in philosophy. Philosophy of psychology: Contemporary readings/edited by Josè Luis Bermudez and Fiona MacPherson* (pp. 407–432). London: Routledge.
Newen, A. (2015). *Understanding others: The person model theory*. Advance online publication. https://doi.org/10.15502/9783958570320
Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science (New York, N.Y.), 308*(5719), 255–258. https://doi.org/10.1126/science.1107621
Oyama, S. (1985/2000). *The ontogeny of information*. Duke University Press. https://doi.org/10.1215/9780822380665
Perez-Zapata, D., Slaughter, V. P., & Henry, J. D. (2016). Cultural effects on mindreading. *Cognition, 146*, 410–414. https://doi.org/10.1016/j.cognition.2015.10.018
Perner, J., Priewasser, B., & Roessler, J. (2018). The practical other: Teleology and its development. *Interdisciplinary Science Reviews, 43*(2), 99–114. https://doi.org/10.1080/03080188.2018.1453246
Piaget, J. (1954). *The construction of reality in the child* (Vol. 20). London: International library of psychology. https://doi.org/10.1016/B978-008044490-1/50022-3 [Place of publication not identified]: Routledge; [publisher not identified].
Piaget, J. (1971). *Genetic epistemology*. New York: Norton.
Piaget, J. (1999) [1951]). *Play, dreams and imitation in childhood. International library of psychology* (Vol. 87). London: Routledge.
Piaget, J., & Inhelder, B. (2000). *The psychology of the child* [1969]. New York: Basic Books.
Polk, D. (1982). *A book called hiraeth: Longing for Wales/compiled by dora polk*. Port Talbot: Alun.
Poulin-Dubois, D., Rakoczy, H., Burnside, K., Crivello, C., Dörrenberg, S., Edwards, K., & Ruffman, T. (2018). Do infants understand false beliefs? We don't know yet – a commentary on Baillargeon, buttelmann and southgate's commentary. *Cognitive Development, 48*, 302–315. https://doi.org/10.1016/j.cogdev.2018.09.005
Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*, 515. https://doi.org/10.1017/S0140525X00076512, 04.
Pyers, J. E., & Senghas, A. (2009). Language promotes false-belief understanding: Evidence from learners of a new sign language. *Psychological Science, 20*(7), 805–812. https://doi.org/10.1111/j.1467-9280.2009.02377.x
Reddy, V. (2018). Why engagement? In A. Newen, L. de Bruin, & S. Gallagher (Eds.), *Oxford library of psychology. The Oxford handbook of 4E cognition*. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198735410.013.23.
Reddy, V., Markova, G., & Wallot, S. (2013). Anticipatory adjustments to being picked up in infancy. *PloS One, 8*(6), Article e65289. https://doi.org/10.1371/journal.pone.0065289
Roby, E., & Scott, R. M. (2018). The relationship between parental mental-state language and 2.5-year-olds' performance on a nontraditional false-belief task. *Cognition, 180*, 10–23. https://doi.org/10.1016/j.cognition.2018.06.017
Rucińska, Z. (2020). Affordances in dennett's from bacteria to bach and back. *AVANT. The Journal of the Philosophical-Interdisciplinary Vanguard, 11*(2). https://doi.org/10.26913/avant.2020.02.05
Ruffman, T., Puri, A., Galloway, O., Su, J., & Taumoepeau, M. (2018). Variety in parental use of "want" relates to subsequent growth in children's theory of mind. *Developmental Psychology, 54*(4), 677–688. https://doi.org/10.1037/dev0000459
Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children's and mothers' mental state language and theory-of-mind understanding. *Child Development, 73*(3), 734–751.
Ruffman, T., Slade, L., Rowlandson, K., Rumsey, C., & Garnham, A. (2003). How language relates to belief, desire, and emotion understanding. *Cognitive Development, 18*(2), 139–158. https://doi.org/10.1016/S0885-2014(03)00002-9
Salo, V. C., Rowe, M. L., Leech, K. A., & Cabrera, N. J. (2016). Low-income fathers' speech to toddlers during book reading versus toy play. *Journal of Child Language, 43*(6), 1385–1399. https://doi.org/10.1017/S0305000915000550
Samuels, R. (2004). Innateness in cognitive science. *Trends in Cognitive Sciences, 8*(3), 136–141. https://doi.org/10.1016/j.tics.2004.01.010
Savitsky, K., Keysar, B., Epley, N., Carter, T., & Swanson, A. (2011). The closeness-communication bias: Increased egocentrism among friends versus strangers. *Journal of Experimental Social Psychology, 47*(1), 269–273. https://doi.org/10.1016/j.jesp.2010.09.005

Schelling, T. C. (1997). *Strategy of conflict* [1963] (2. edition). London: Harvard University, 16. printing.
Scott, R. M., & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences, 21*(4), 237–249. https://doi.org/10.1016/j.tics.2017.01.012
Spaulding, S. (2018). How we understand others: Philosophy and social cognition. In *Routledge focus on philosophy* (1 edition). New York: Routledge.
Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science, 10*(1), 89–96. https://doi.org/10.1111/j.1467-7687.2007.00569.x
Stone, J. E., Carpendale, J. I. M., Sugarman, J., & Martin, J. (2012). A Meadian account of social understanding: Taking a non-mentalistic approach to infant and verbal false belief understanding. *New Ideas in Psychology, 30*(2), 166–178. https://doi.org/10.1016/j.newideapsych.2011.09.002
Symons, D. K., Peterson, C. C., Slaughter, V., Roche, J., & Doyle, E. (2005). Theory of mind and mental state discourse during book reading and story-telling tasks. *British Journal of Developmental Psychology, 23*(1), 81–102. https://doi.org/10.1348/026151004X21080
Taumoepeau, M. (2019). Culture, communication and socio-cognitive development: Understanding the minds of others. In T. Tulviste, D. L. Best, & J. L. Gibbons (Eds.), *Children's social worlds in cultural context* (Vol. 78, pp. 41–54). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-27033-9_4.
Tompkins, V., Benigno, J. P., Kiger Lee, B., & Wright, B. M. (2018). The relation between parents' mental state talk and children's social understanding: A meta-analysis. *Social Development, 78*, 1052. https://doi.org/10.1111/sode.12280
Turner, S. P. (1994). *The social theory of practices: Tradition, tacit knowledge and presuppositions/Stephen Turner*. Cambridge: Polity Press.
Turner, S. P. (2018). *Cognitive science and the social: A primer/stephen P. Turner (1st)*. London: Routledge.
Vandell, D. L., & Wilson, K. S. (1987). Infants' interactions with mother, sibling, and peer: Contrasts and relations between interaction systems. *Child Development, 58*(1), 176. https://doi.org/10.2307/1130299
Vinden, P. G. (1999). Children's understanding of mind and emotion: A multi-culture study. *Cognition & Emotion, 13*(1), 19–48. https://doi.org/10.1080/026999399379357
Vinden, P. G. (2002). Understanding minds and evidence for belief: A study of mofu children in Cameroon. *International Journal of Behavioral Development, 26*(5), 445–452. https://doi.org/10.1080/01650250143000391
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes/L. S. Vygotsky ; edited by Michael Cole … [et al.] ; [translated from the Russian]. Cambridge, Mass.* London: Harvard University Press.
Ward, D., Silverman, D., & Villalobos, M. (2017). Introduction: The varieties of enactivism. *Topoi, 36*(3), 365–375. https://doi.org/10.1007/s11245-017-9484-6
Waxman, S., Fu, X., Arunachalam, S., Leddon, E., Geraghty, K., & Song, H.-j. (2013). Are nouns learned before verbs? Infants provide insight into a longstanding debate. *Child Development Perspectives, 7*(3). https://doi.org/10.1111/cdep.12032
Weekes, D. P., Kagan, S. H., James, K., & Seboni, N. (1993). The phenomenon of hand holding as a coping strategy in adolescents experiencing treatment-related pain. *Journal of Pediatric Oncology Nursing : Official Journal of the Association of Pediatric Oncology Nurses, 10*(1), 19–25. https://doi.org/10.1177/104345429301000105
Wellman, H. M. (2014). *Making minds: How theory of mind develops. Oxford series in cognitive development*. Oxford, New York: Oxford University Press.
Wellman, H. M. (2018). Theory of mind: The state of the art. *European Journal of Developmental Psychology, 7*, 1–28. https://doi.org/10.1080/17405629.2018.1435413
Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of theory-of-mind understandings in Chinese children. *Psychological Science, 17*(12), 1075–1081. https://doi.org/10.1111/j.1467-9280.2006.01830.x
Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x
Westra, E. (2017). Pragmatic development and the false belief task. *Review of Philosophy and Psychology, 8*(2), 235–257. https://doi.org/10.1007/s13164-016-0320-5
Westra, E., & Carruthers, P. (2017). Pragmatic development explains the theory-of-mind scale. *Cognition, 158*, 165–176. https://doi.org/10.1016/j.cognition.2016.10.021
Westra, E., & Schönherr, J. (2019). Beyond 'interaction': How to understand social effects on social cognition. *The British Journal for the Philosophy of Science, 70*(1), 27–52. https://doi.org/10.1093/bjps/axx041
Wierzbicka, A. (1999). *Emotions across languages and cultures: Diversity and universals/Anna Wierzbicka. Studies in emotion and social interaction* (Second series). Cambridge: Cambridge University Press.
Wierzbicka, A. (2005). Empirical universals of language as a basis for the study of other human universals and as a tool for exploring cross-cultural differences. *Ethos, 33*(2), 256–291. https://doi.org/10.1525/eth.2005.33.2.256
Wierzbicka, A. (2006). On folk conceptions of mind, agency and morality. *Journal of Cognition and Culture, 6*(1), 165–179. https://doi.org/10.1163/156853706776931286
Applications of cognitive linguistics. In Yu, N. (Ed.), *The Chinese heart in a cognitive perspective: Culture, body, and language* (Vol. 12), (2009). Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110213348. Retrieved from http://reference-global.com/doi/book/10.1515/9783110213348.
Zawidzki, T. W. (2013). *Mindshaping: A new framework for understanding human social cognition*. Cambridge, MA: MIT Press.
Zittoun, T., & Gillespie, A. (2015). Internalization: How culture becomes mind. *Culture & Psychology, 21*(4), 477–491. https://doi.org/10.1177/1354067X15615809
Ziv, M., Smadja, M.-L., & Aram, D. (2014). Mothers' and teachers' mental-state discourse with preschoolers during storybook reading. *Journal of Cognitive Education and Psychology, 13*(1), 103–119. https://doi.org/10.1891/1945-8959.13.1.103

PSYCHOLOGIA ROZWOJOWA, 2019 ∗ tom 24, nr 2, s. 15–28
doi:10.4467/20843879PR.19.007.10890
www.ejournals.eu/Psychologia-Rozwojowa

ROBERT MIRSKI orcid.org/0000-0002-1608-4734

Instytut Filozofii, Katolicki Uniwersytet Lubelski Jana Pawła II
Faculty of Philosophy, John Paul II Catholic University of Lublin
e-mail: robertmirski@kul.lublin.pl

Krytyka natywizmu jawnego i ukrytego w badaniach nad dziecięcymi teoriami umysłu

A Critique of Explicit and Implicit Nativism in Research on Children's Theories of Mind

Abstract. The traditional theories of theory-of-mind development – modularist nativism, theory theory, and the two-systems theory – share a common model of mental representation. According to that model, the normative content of representation is encoded in its physical vehicle. In the present article, I point out that this claim entails the view that representation cannot emerge out of non-representational phenomena. This leads to the need of positing foundational mental content – foundationalism – and viewing cognitive development only as a reconfiguration of the innately given representations. As a result, all three models are forced to claim innate mental content, although only the modular nativists explicitly acknowledge it. Further, the idea that mental content is innate faces its own challenges: nativism does not seem to be a tenable position in either the "biological" or "psychological" sense of the term. I argue that nativism is a symptom of theoretical limitations, not a legitimate division of labor between psychology and other sciences.

Keywords: theory of mind, mindreading, mental representation, nativism, foundationalism, emergence

Słowa kluczowe: teoria umysłu, mindreading, natywizm, emergencja, reprezentacje mentalne, fundacjonizm

WSTĘP

Stanowisko natywistyczne jest wciąż jednym z dominujących poglądów w psychologicznych badaniach nad rozwojem teorii umysłu (*theory of mind*, ToM) (e.g. Carruthers, 2013, 2015; Fodor, 1992; Helming, Strickland, Jacob, 2016; Leslie, Friedman, German, 2004; Onishi, Baillargeon, 2005; Scholl, Leslie, 2001; Scott, Baillargeon, 2017; Westra, 2017; Westra, Carruthers, 2017). Zwolennicy tego stanowiska zakładają, że centralnym mechanizmem poznawczym umożliwiającym poznanie społeczne jest wrodzony moduł czytania w umyśle; wrodzona teoria umysłu, która zawiera pojęcia przekonania i pragnienia dostępne już niemowlętom.

Natywizm wszelkiego rodzaju jest zazwyczaj obiektem krytyki, która stara się wykazać, że doświadczenie jest niezbędne do nabycia pojęć uznawanych za wrodzone lub że wystarcza ono do ich formacji. W badaniach nad poznaniem społecznym są to głównie argumenty powołujące się na wyniki empiryczne, które, jak argumentują przeciwnicy postulatu wrodzoności, są albo niespójne z perspektywą natywizmu, albo niewystarczające, żeby wykluczyć alternatywne interpretacje (Apperly, 2012; Apperly, Butterfill, 2009; Gopnik, 2003; e.g. Gopnik,

Wellman, 1992; Perner, Ruffman, 2005; Ruffman, Taumoepeau, 2014). W niniejszym tekście przyjmę nieco inną perspektywę: pokażę, że natywizm jest również koniecznym stanowiskiem tych bardziej „empirycystycznych" teorii (teorii teorii, teorii dwóch systemów) z powodów czysto teoretycznych – model reprezentacji, zakładanych przez tradycyjne podejścia ToM, wymusza postulat początkowych reprezentacji, z których rozwój poznawczy się rozpoczyna. W rezultacie te stanowiska, pomimo tego, że unikają „natywistycznej" nomenklatury, są również zmuszone do przyjęcia wrodzoności dla postulowanych fundamentów rozwojowych.

Następnie omawiam problem samego postulatu wrodzoności reprezentacji mentalnych: co miałby on oznaczać i czy jest do utrzymania. Rozważam dwa główne znaczenia „wrodzoności" – zapożyczenie tego pojęcia z biologii oraz rozumienie go jako metodologicznej granicy psychologii. Pokazuję, że oprócz ograniczeń teoretycznych wymuszających ten pogląd nie ma powodów, aby z góry zakładać, że wyłonienie się pierwszych reprezentacji mentalnych leży poza obszarem psychologii.

EMPIRYCZNE TŁO

Obecne wyniki badań empirycznych nad poznaniem społecznym można podzielić na trzy kategorie. Przedstawiam je poniżej w celu nakreślenia tła dla następujących po nich rozważań teoretycznych.

(1) Dzieci rozwiązują test fałszywych przekonań w paradygmacie spontanicznej reakcji dużo wcześniej niż tradycyjny test fałszywych przekonań (*false-belief task*, FBT), który wymaga odpowiedzi dziecka na pytanie testowe[1]. Ten pierwszy potrafią rozwiązać już piętnastomiesięczne dzieci, natomiast z tradycyjną wersją zaczynają sobie radzić dopiero czterolatki i starsze dzieci (Kovács, Téglás, Endress, 2010; Onishi, Baillargeon, 2005; Poulin-Dubois, Chow, 2009; Scott, Baillargeon, 2009; Scott, Baillargeon, Song, Leslie, 2010; Träuble, Marinović, Pauen, 2010).

(2) Czynniki językowe oraz społeczne w kontekście rozwoju dziecka wpływają na rozwój poznania społecznego, w tym na wiek, w którym dzieci zdają tradycyjny FBT (e.g. Kristen, Sodian, 2014; Milligan, Astington, La Dack, 2007; Nelson, 2005; Ruffman, Slade, Rowlandson, Rumsey, Garnham, 2003; de Villiers, de Villiers, 2014).

(3) Istnieją kultury, w których tłumaczy się zachowanie innych ludzi na sposoby radykalnie odmienne od zachodniej psychologii potocznej opartej na pojęciach umysłu i stanów mentalnych. Ponadto dzieci z tych kultur zazwyczaj rozwiązują testy poznania społecznego w innym wieku niż dzieci pochodzące z kultur Zachodu (e.g. Dixson, Komugabe-Dixson, Dixson, Low, 2017; Mayer, Träuble, 2012; Mills, 2001; Strijbos, De Bruin, 2013; Vinden, 1996; Wellman, Fang, Liu, Zhu, Liu, 2006; Wellman, Fang, Peterson, 2011).

Punkty 2 i 3 sugerują, że zdolność do rozumienia innych umysłów jest w znaczącym stopniu modulowana przez czynniki środowiskowe, takie jak kultura, język czy kontekst rodzinny. Wyniki wspomniane w punkcie 1 sugerują jednak istnienie jakiejś zdolności pozwalającej na przynajmniej minimalną kompetencję społeczną już w wieku 15 miesięcy. Empirycystyczno-natywistyczne „wahadło" kontynuuje swój nieustający ruch również na tle tych wyników (Allen, Bickhard, 2013): modularni natywiści postulują, że wyniki z punktu 1 są dowodem istnienia wrodzonego modułu czytania w umyśle opartego na pojęciach przekonania i pragnienia, podczas gdy zwolennicy poglądów bardziej empirycystycznych argumentują, że wcale tak być nie musi – dużo prostszy mechanizm może wyjaśnić te wyniki, a teoria umysłu oparta na zrozumieniu przekonań i pragnień jest wynikiem doświadczenia w środowisku, co współgra z wynikami wspomnianymi w punktach 2 i 3.

Poniżej zwracam uwagę, że oba te stanowiska, pomimo różnic, które je dzielą, są przedmiotem tych samych teoretycznych ograniczeń. Zwracam uwagę na fakt, że natywizm jest poglądem wymuszonym przez tradycyjny pogląd na reprezentację mentalną. Pogląd ten jest wspólny zarówno stanowiskom jawnie natywistycznym, jak i tym, które unikają tego terminu. Zgodnie z prezentowanym tutaj argumentem stanowiska, które trwają przy korespondencyjnym modelu reprezentacji, są zmuszone ze względów teo-

retycznych do przyjęcia postulatu pierwotnych reprezentacji. Różnica pomiędzy stanowiskami natywistycznymi i tymi, które się za takie nie uważają (lecz wciąż używają korespondencyjnej reprezentacji), jest jedynie taka, że te drugie przemilczają problem genezy pierwszych reprezentacji. Przedstawiony argument czerpie z krytyki prowadzonej przez Marka Bickharda od kilkudziesięciu lat, gdzie można dostrzec szerszą skalę tego problemu niż jedynie dyskutowane tutaj badania nad ToM (Allen, Bickhard, 2013; Bickhard, 2001, 2015, 2016; Bickhard, Richie, 1983; Bickhard, Terveen, 1995; Campbell, Bickhard, 1986).

NATYWIZM – POGLĄD BARDZIEJ POWSZECHNY NIŻ SIĘ WYDAJE

Tradycyjne stanowiska ToM (teoria teorii, modularny natywizm oraz niedawno zaproponowana teoria dwóch systemów) pracują z zasadniczo tym samym modelem reprezentacji mentalnej: treść semantyczna – czyli normatywność reprezentacji, to, co sprawia, że może być ona prawdziwa lub nie – jest rozumiana jako wynikająca z korespondencji z tym, co jest reprezentowane. Reprezentacja „koduje" informację na temat desygnatu, stąd ten ogólny pogląd Bickhard określa mianem enkodyzmu (*encodingism*). Pogląd ten jest standardowy nie tylko w badaniach nad teorią umysłu, lecz także w „podręcznikowej" psychologii poznawczej (zob. Nęcka, Orzechowski, Szymura, 2006). Fakt ten, niezależnie od bardziej szczegółowych różnic poszczególnych teorii, narzuca wspólne tym podejściom ograniczenia teoretyczne, które prowadzą do jawnego lub niejawnego natywizmu reprezentacji.

Korespondencyjny lub „enkodystyczny" model reprezentacji ma swoje korzenie w klasycznej teorii umysłu obliczeniowego, który ujmował poznanie za pomocą metafory komputera. Energia płynąca ze środowiska ma być przetworzona (*transduced*) na format układu nerwowego. Podstawowym założeniem – i centralnym problemem – jest tutaj to, że owo przetworzenie ma skutkować powstaniem normatywnej treści mówiącej coś o środowisku. W klasycznych natywistycznych podejściach ta normatywność zawierała się w pojęciach, swoistym „języku myśli" (Fodor, 1975), natomiast w podejściach bardziej empirycystycznych miała być ona zawarta w dużo prostszych reprezentacjach sensorycznych. Jednak w obu przypadkach konieczny jest postulat, że to, co czysto przyczynowe – energia docierająca do systemu poznawczego – zostaje przemianowane na to, co normatywne, co przynajmniej w minimalnym stopniu „orzeka" coś o rzeczywistości. Jest to konieczne, ponieważ na tym właśnie polega poznanie – na zdobywaniu *prawdziwej* informacji o rzeczywistości i wykorzystywaniu jej w adaptacyjnych zachowaniach.

Na przykład sama energia elektromagnetyczna światła nie mówi nic o otaczającej rzeczywistości – system poznawczy musi już „wiedzieć", że pewne zmienności w świetle odnoszą się do przedmiotów czy bardziej ogólnych warunków mających miejsce w rzeczywistości. Tylko wtedy możemy otrzymać normatywną treść, na podstawie której system może podjąć jakieś działanie i spodziewać się sukcesu.

Staje się więc jasne, że sama korespondencja, na której ma zasadzać się treść semantyczna, nie jest wystarczającym warunkiem jej istnienia. Reakcja układu wzrokowego koresponduje zarówno z geometrią otoczenia, jak i samym światłem oraz także z molekularną strukturą przedmiotów i początkami wszechświata – postulat, że jeden zakres z tej korespondencji jest tym, o czym reprezentacja orzeka, jest czysto arbitralny. Sam fakt istnienia informacji nie wyjaśnia, w jaki sposób system o niej wie.

Centralny problem polega na tym, że pomimo długoletnich starań w ramach korespondencyjnego modelu reprezentacji nie udało się nikomu wyjaśnić, w jaki sposób to, co czysto przyczynowe, staje się normatywne, kiedy dociera do umysłu. W konsekwencji psychologia poznawcza poprzestała na założeniu, że treść reprezentacji jest już obecna w umyśle: bodźce ze środowiska aktywują schematy normatywne już w nim zawarte. Najbardziej wyczerpującą analizę tego problemu zaoferował Jerry A. Fodor, który podążając logicznymi konsekwencjami przyjętego modelu reprezentacji, został zmuszony do postulatu wrodzoności praktycznie wszystkich mentalnych reprezentacji, co wbrew

temu, co się często o nim orzeka, nie było stanowiskiem, które uważał za właściwe:

> (…) ten argument nie może być poprawny, (…) natywizm, który w rezultacie otrzymujemy, jest nie do utrzymania, (…) coś bardzo ważnego musiało nam umknąć. Wydaje mi się, że wykazane konsekwencje nie stanowią apriorycznego argumentu za natywizmem, a raczej wskazują na to, że musi istnieć jakiś model uczenia się, który jest tak radykalnie różny od tego, co do tej pory sobie wyobrażaliśmy, że sam nie jestem w stanie zasugerować, jak mógłby on wyglądać (wypowiedź Fodora w Piattelli-Palmarini, 1980, s. 269)[2].

Problem wrodzoności reprezentacji nie jest nigdzie indziej tak problematyczny jak w poznawczej psychologii rozwojowej – zakładając, że to, co mentalne, już istnieje, przyznajemy w gruncie rzeczy, że nie jesteśmy w stanie odpowiedzieć na podstawowe pytanie tej nauki – skąd bierze się umysł. Presuponujemy to, co powinno być wyjaśnione. W kontekście badań rozwojowych problem ten został określony mianem fundacjonizmu (*foundationalism*) (Allen, Bickhard, 2013; Bickhard, Terveen, 1995) – przyjęty model reprezentacji jako korespondencji wymusza założenie istnienia początkowych reprezentacji, ponieważ jest logicznie niemożliwe wyprowadzenie normatywności z przyczynowości.

Kiedy już przyjmiemy, że normatywność istnieje w postaci wrodzonych reprezentacji, to modelowanie rozwoju poznawczego staje się relatywnie bezproblematyczne. Treść fundacyjnych reprezentacji jest łączona na wiele sposobów czysto mechanistycznie, co prowadzi do utworzenia złożonych pojęć i pozornego rozwoju poznawczego. Trzeba jednak zauważyć, że nie rozwiązaliśmy tutaj podstawowego problemu; nie powstaje w ten sposób zasadniczo nowa treść, a jedynie „opracowywane" są proste reprezentacje.

FUNDACJONIZM W TEORIACH ToM

W rezultacie powyższych ograniczeń teoretycznych każdy model, który przyjmuje, że reprezentacja „koduje" swoją treść, musi założyć istnienie podstawowych reprezentacji, czy to w postaci pojęć, czy „prostszych" struktur normatywnych. Problem ten jest widoczny w podejściach ToM, zarówno tych „natywistycznych", jak i tych, które unikają tego terminu.

Modularny natywizm ToM otwarcie przyjmuje konsekwencje swojego modelu poprzez postulat, że pierwsze reprezentacje są wrodzone. Podobnie jak w przypadku modularnych modeli innych zdolności poznawczych (Lightfoot, 1989; Pinker, 2014/1994; Wynn, 1992), mechanizm czytania w umyśle ma wyglądać następująco: istnieje genetycznie determinowany mechanizm lub moduł, który jest przeznaczony do konkretnej dziedziny poznawczej – w naszym wypadku czytania w umyśle. Mechanizm ten jest niezależny od reszty systemu poznawczego i niewrażliwy na jakiekolwiek pozasystemowe czynniki rozwojowe. Rozwija się według genetycznie (lub „biologicznie") predeterminowanego grafiku (*timetable*). Informacja, która jest wrodzona w jego postaci, zawiera podstawowe pojęcia (przekonanie, pragnienie, widzenie czy udawanie) oraz heurystyki (np. „jeśli ktoś coś zobaczy, to tworzy na ten temat przekonanie"), co umożliwia dziecku selekcję relewantnej informacji sensorycznej, przetłumaczenie jej na format kodu modułu oraz wyciągnięcie podświadomych wniosków. Wnioski te są następnie inkorporowane w centralny system, co dodaje aspekt psychiki drugiego człowieka w postrzeganej rzeczywistości. Jak twierdzą Evan Westra i Peter Carruthers (2017), wrodzony moduł czytania w umyśle może być wzbogacony: albo w jakiś sposób jego wyjściowy zestaw pojęć się rozszerza (Westra, Carruthers, 2017), albo harmonizacja z resztą systemu poprawia się, co umożliwia bardziej złożone wnioskowania wynikające ze współpracy modułu z innymi modułami/systemem centralnym (Carruthers, 2015; Fodor, 1992; Leslie, Friedman, German, 2004).

Zwolennicy natywizmu tłumaczą różnice rozwojowe z punktów 2 i 3 przedstawionych na początku tekstu oraz wpływ czynników środowiskowych w nich wspomniany, przytaczając inne problemy niż brak samego zrozumienia umysłu przez dziecko. Twierdzą, że dany test albo wymaga pojęć bardziej złożonych niż moduł czytania w umyśle dostarcza (np. zazdrości), albo

trudności dziecka tłumaczą się klasycznym już rozbratem pomiędzy kompetencją a czynnikami wykonawczymi (*competence-performance*). To ta druga ścieżka jest konieczna przy interpretacji wyników FBT i podobnych testów: testy te badają zdolności, które natywiści uznają za wrodzone (jak zrozumienie fałszywych przekonań), a dzieci wciąż rozwiązują je w stopniu zależnym od kontekstu rozwoju.

Zwolennicy natywizmu postulują, że dziecko rozumie fałszywe przekonania od urodzenia, a różnice w rozwiązywaniu przez nie testów ToM skorelowane z czynnikami środowiskowymi mają swe źródło w (a) niezrozumieniu implikatur pytań testowych, (b) braku odpowiedniego słownictwa w języku, w którym dziecko dorasta, lub (c) nierozwiniętych jeszcze dostatecznie funkcjach wykonawczych albo ogólnych procesach obliczeniowych (Fodor, 1992; cf. Helming i in., 2016; Westra, Carruthers, 2017). Innymi słowy, czynniki środowiskowe wpływają na rozwój zdolności komunikacyjnych dziecka i funkcji wykonawczych, nie na rozwój jego modułu czytania w umyśle. To te pierwsze stoją na drodze ekspresji zrozumienia innych umysłów przez początkowe lata życia dziecka, osiągając wystarczające stadium rozwoju w różnym czasie, w zależności od kontekstu rozwojowego.

Natywiści mówią otwarcie o wrodzonych pojęciach. Jednak również stanowisko racjonalnego konstruktywizmu (teoria teorii), które jest głównym przeciwnikiem modularnego natywizmu, pada ofiarą fundacjonizmu i jest zmuszone do przynajmniej implicytnego zakładania wrodzoności (Gopnik, 2009, 2010, 2011; Gopnik, Meltzoff, Kuhl, 1999; Gopnik, Wellman, 1992, 2012; Wellman, 2014). Teoria teorii zakłada, że rozwój poznawczy zachodzi analogicznie do rozwoju teorii w nauce, i jej zwolennicy posługują się metaforą „małego naukowca", którym ma być dziecko. Ma zacząć ono swój rozwój poznawczy z zestawem pojęć służących mu do stawiania hipotez, które potem kontrastuje z rzeczywistością poprzez przeprowadzanie „eksperymentów". Proces ten jest czysto formalny i jest wsparty teorią wnioskowań Bayesowskich (Gopnik, Bonawitz, 2015; Gopnik, Tenenbaum, 2007). Nie wyjaśnia to jednak, skąd reprezentacje używane w tym procesie zdobywają swoją treść, i w rezultacie mamy do czynienia z kolejnym przykładem fundacjonizmu. Fakt ten widać klarownie w poniższym fragmencie z książki Henry'ego M. Wellmana:

> Każdy model rozwoju musi ustalić zarówno stany początkowe organizmu, jak i mechanizmy rozwoju – początki oraz rozwój. Nie sposób obyć się bez tych dwóch elementów, aczkolwiek oczywiście można postulować bogatsze lub mniej zasobne stany początkowe oraz bardziej złożony lub prostszy rozwój. Każdy model, taki jak teoria teorii, który podkreśla rolę uczenia się w rozwoju, musi zawierać stany początkowe oraz mechanizmy uczenia się, które przekształcają stany początkowe w późniejsze struktury (Wellman, 2014, s. 197).

Rozwój rozumiany jako rozwój teorii musi zacząć się od postulatu istniejących już pojęć – pierwszej minimalnej teorii. Żeby w ogóle się rozpoczął, musi zostać postawiona pierwsza hipoteza, a żeby to nastąpiło, musi już istnieć teoria i jej pojęcia. W konsekwencji, pomimo że Alison Gopnik, Wellman i ich współpracownicy unikają pojęcia wrodzoności, wydaje się, że nie mają innego wyjścia, jak również uciec się do natywizmu. W każdym razie teoria teorii nie jest odpowiednim modelem rozwoju poznawczego, ponieważ presuponuje to, co taki model wyjaśnić powinien – emergencję normatywności.

Nowszym wariantem podejść ToM jest teoria dwóch systemów (Apperly, 2012; Apperly, Butterfill, 2009; Butterfill, Apperly, 2013; Low, Apperly, Butterfill, Rakoczy, 2016). Zakłada się tutaj, że ludzie są w posiadaniu dwóch systemów, którymi posługują się w nawigacji życia społecznego. System 1 jest prostym mechanizmem, który umożliwia „śledzenie" stanów mentalnych innych ludzi, ale nie reprezentuje ich w sposób refleksyjny (czyli jako stanów mentalnych). System 2 jest świadomym mechanizmem refleksji nad życiem mentalnym drugiego człowieka.

Motywacja do postulatu dwóch systemów czytania w umyśle była empiryczna: przepaść pomiędzy 15-miesięcznymi dziećmi zdającymi test fałszywych przekonań w paradygmacie

spontanicznej reakcji a dopiero 4-latkami zdającymi eksplicytny test fałszywych przekonań jest tutaj wyjaśniona poprzez funkcjonowanie dwóch systemów poznawczych, z czego drugi rozwija się później. Ian A. Apperly i Stephen A. Butterfill przeprowadzili wiele dalszych badań, które pokazują, że nawet u dorosłych oba te systemy działają niezależnie (Apperly, 2012).

W kontekście problemu fundacjonizmu teoria dwóch systemów nadal napotyka te same aporie, ponieważ jej autorzy również presuponują korespondencyjną teorię reprezentacji. Apperly i Butterfill są dużo ostrożniejsi w swoich teoretycznych postulatach niż zwolennicy dwóch poprzednich podejść, jednak wciąż nawet ich system 1 rozpoczyna swoje istnienie z zestawem reprezentacji:

> Nie mamy na celu argumentowanie, że osoba może śledzić przekonania, prawdziwe i fałszywe, bez jakiegokolwiek poznania opartego na teorii umysłu. Chcemy raczej zaproponować minimalną formę poznania opartego na teorii umysłu. (…) minimalna teoria umysłu reprezentuje stany mentalne podobne do przekonań, ale nie reprezentuje przekonań oraz innych postaw propozycjonalnych jako takich (Butterfill, Apperly, 2013, s. 3).

Teoria dwóch systemów wciąż skupia się na problemie askrypcji stanów mentalnych obserwowanym podmiotom. W konsekwencji system 1, pomimo swojej minimalistycznej struktury, wciąż presuponuje, że dziecko jest w stanie reprezentować podmioty i przedmioty oraz że podmioty rejestrują przedmioty. Czyli znowu mamy do czynienia z założeniem normatywności poznania bez jej wyjaśnienia. Autorzy dwóch systemów, podobnie jak teoretycy teorii, unikają twierdzeń na temat wrodzoności, lecz nie sposób nie odnieść wrażenia, że i oni nie mają innego wyjścia.

Przedstawione teorie ToM padają ofiarą fundacjonizmu ze względu na model reprezentacji, który zakładają. W tym kontekście ich główną różnicą jest to, jak złożone reprezentacje są początkiem rozwoju (np. dla Wellmana jest to pojęcie pragnienia i świadomości (*awareness*), dla modularystów pojęcia przekonania i pragnienia) – jednak w każdym z trzech przypadków rozwój normatywności jako takiej pozostaje niewyjaśniony.

PROBLEMY Z POJĘCIEM WRODZONOŚCI

Z racji tego, że wrodzoność pierwszych reprezentacji jest jawnym (natywizm) i potencjalnym (teoria teorii, teoria dwóch systemów) wyjściem z fundacjonizmu, poniżej rozważam, czy sam postulat wrodzoności ma sens.

Wrodzoność z perspektywy biologii

Timothy P. Racine (2013) zauważa, że podejścia zakładające wrodzoną wiedzę zazwyczaj wykorzystują neodarwinowski pogląd na wrodzoność, że wrodzone pojęcia były obiektem presji selekcyjnych w filogenezie ze względu na ich adaptacyjność i stąd są one zakodowane w genach i przez to koniecznie rozwijają się w ontogenezie (zob. np. Carruthers, 2013, s. 151). Problem fundacjonizmu nie jest więc rozwiązany, a jedynie zrzucony na biologię. Biologia jednak podważa tak prostolinijne stanowisko; staje się jasne, że złożone cechy fenotypiczne, takie jak pojęcia, są wynikiem interakcji czynników rozwojowych, która wymaga wyjaśnienia, a nie przyjęcia za pewnik.

W biologii panuje obecnie atmosfera wielkich zmian, a niektórzy okrzyknęli już wiek XXI wiekiem biologii (Venter, Cohen, 2004). Jednym z głównych problemów w tym rewolucyjnym klimacie jest pojęcie wrodzoności. Pod wpływem prac takich naukowców jak Richard Ch. Lewontin czy Stephen J. Gould współczesna biologia dużo ostrożniej postuluje adaptacjonistyczne wyjaśnienia rozwoju cech fenotypicznych; pogląd, że są wrodzone i konieczne w ontogenezie (Gould, Lewontin, 1979; Lewontin, 2001; Oyama, 1985/2000).

Badania z zakresu biologii pokazują, że rozwój osobniczy jest determinowany wieloma czynnikami i zachodzącymi pomiędzy nimi relacjami (*multiply contigent*) (Elman, 1996; Gould, Lewontin, 1979; Gould, Vrba, 1982; Mameli, Bateson, 2011; Oyama, 1985/2000; Pigliucci, Müller, 2010). Wielu psychologów, szczególnie rozwojowych, ponagla, że musi być to uwzględnione w badaniach nad rozwojem poznawczym (Carpendale, Hammond, Atwood, 2013, s. 130; Carpendale, Wereha, 2013, s. 208;

Lewis, Carpendale, Stack, 2013, s. 159–160; Lewkowicz, 2011; Spencer i in., 2009). Zwracają oni uwagę na to, że istnieje wiele elementów, których interakcja prowadzi do wyłonienia się form poznawczych, i stąd jakakolwiek rozmowa o „wrodzoności", czyli o zakodowaniu w genach czy jakiekolwiek innej „preegzystencji", wypacza rzeczywisty sposób, w jaki geny lub inne interaktanty rozwojowe mają znaczenie w ontogenezie. „Interakcjonistyczna lekcja", którą się wynosi ze współczesnej biologii, jest taka, że geny mają dane znaczenie dla ontogenezy tylko i wyłącznie w kontekście innych, zarówno wewnątrz-, jak i zewnątrzorganizmowych czynników/interaktantów. Innymi słowy, mają one swoją „informację" o konkretnej cesze tylko wtedy, kiedy inne przyczyny występują. Ta interakcyjna natura zjawisk ontogenetycznych sprawia, że nie ma zbyt wielkiego sensu mówić o wrodzoności jako o cechach „zapisanych w genach". Można by równie dobrze powiedzieć, że wrodzone cechy są „zapisane w środowisku", jako że moc przyczynowa genów jest możliwa dzięki konkretnemu zakresowi środowisk: „Geny są niezbędnym czynnikiem w systemie wzajemnie oddziałujących na siebie czynników; złożoność wyłania się, a nie preegzystuje na jakimkolwiek wcześniejszym etapie rozwoju" (Carpendale, Wereha, 2013, s. 208).

Natywistyczny pogląd na rozwój poznania społecznego ignoruje powyższy postęp w zrozumieniu rozwoju poznawczego i używa pojęcia wrodzoności, które nie jest spójne ze współczesną biologią. Spójrzmy na poniższy cytat z Petera Carruthersa (2013):

> Hipoteza wczesnego czytania w umyśle postuluje wrodzoną wiedzę podstawową lub wrodzenie ustrukturalizowany mechanizm obliczeniowy (albo obydwa powyższe), które przypominają prostą teorię umysłu. Ciężar wyjaśnień pada więc na procesy ewolucyjne: trzeba wykazać, że zaszły wystarczające presje adaptacyjne w życiu naszych przodków, aby mechanizm czytania w umyśle wyewoluował. Istnieje już pokaźna grupa badań, które sugerują, że właśnie tak było. Zysk, jaki organizm czerpie z takiego mechanizmu, bierze się stąd, że umożliwia on tak zwaną inteligencję machiavelistyczną (Byrne and Whiten, 1988, 1997) czy typowo ludzkie formy kooperacji i kolaboracji (Richerson and Boyd, 2005; Hrdy, 2009), czy jakąkolwiek kombinację powyższych (Carruthers, 2013, s. 151).

Według Carruthersa głównym wyzwaniem natywizmu jest przedstawienie adaptacyjnych argumentów *z filogenezy*, które domniemanie mają wyjaśniać *ontogenetyczny rozwój* danej cechy. W rzeczywistości adaptacyjne wyjaśnienia nie są mocnymi wyjaśnieniami rozwoju osobniczego; filogenetyczne adaptacje *nie są* preformowanymi fenotypami, które każdy osobnik danego gatunku koniecznie wykształca w ontogenezie. Mowa o „wrodzonych cechach" ma miejsce w analizach filogenetycznych, gdzie termin oznacza „zazwyczaj obecny w osobnikach danego gatunku w danym środowisku"; analizach, gdzie przyjmuje się, że środowiskowe kontyngencje są niezmienne, i skupia się uwagę na zmianach w fenotypie na przestrzeni czasu filogenetycznego. Kiedy jednak jesteśmy zainteresowani ontogenezą, staramy się wyjaśnić dokładnie to, co neodarwinizm uogólnia – kontyngencje rozwoju, i to, jak spuścizna filogenetyczna wchodzi w interakcję z konkretnym kontekstem rozwojowym. Fakt, że dana cecha wyewoluowała przez adaptację w filogenezie, nie mówi nam praktycznie nic o tym, w jaki sposób wyłania się ona w ontogenezie (Oyama, 1985/2000, s. 25).

W tym kontekście trzeba pamiętać, że selekcja genetyczna zachodzi w środowisku i konkretne elementy tego środowiska są niezbędne, żeby wyselekcjonowane geny wykształciły w ontogenezie cechę, przez którą zostały wyselekcjonowane. To, co jest więc „wrodzone", to nie genetyczne zakodowanie, które jest problemem biologa, a raczej stabilność pomiędzy organizmem i środowiskiem, którą każdy model rozwoju musi opisać i wyjaśnić.

Wrodzoność jako psychologiczne podstawy

W minimalnym sensie fundacjonizm prowadzi do stanowiska, w którym badacz stwierdza, że reprezentacje stanowiące początek rozwoju nie są jego dziedziną – wyjaśnienie ich emergencji ma leżeć poza obszarem jego zainteresowania. Richard Samuels (2002, 2004), po przeglądzie

dostępnych znaczeń pojęcia wrodzoności, dochodzi do podobnego wniosku, że w najbardziej obiecującym znaczeniu wrodzoności w psychologii i kognitywistyce nie używa się pojęcia wrodzoności zapożyczonego z biologii (gdzie, jak pokazaliśmy, ono tak naprawdę nie funkcjonuje w obecnych czasach), ale raczej ma specyficzne dla tych nauk znaczenie. Według tego poglądu wrodzone struktury poznawcze to takie, których psychologia nie może wyjaśnić. Jej wyjaśnienie miałoby spaść na inną, „niższą" naukę i jak sam Samuels sugeruje (i wydaje się, że wyjścia tutaj innego za bardzo nie ma) tą nauką ma być biologia.

Stwierdzenie jednak, że psycholog zajmujący się rozwojem poznawczym miałby postrzegać problem rozwoju pierwszych pojęć jako „nie swoją działkę", wydaje się co najmniej zaskakujące. Analogicznie, trudno znaleźć biologa, który by twierdził, że problem początków życia jest domeną wyłącznie „niższych" nauk, a biologia nie ma w tej kwestii nic do powiedzenia. W obliczu powyższych rozważań można jednak zrozumieć, dlaczego tak trudno zrezygnować z pojęcia wrodzoności w naukach o umyśle – dominująca w nich teoria wymusza takie stanowisko i nie ma po prostu innego wyjścia. Można jedynie przemilczeć genezę pierwszych reprezentacji, co w gruncie rzeczy nie stanowi żadnej alternatywy.

Jak Samuels zauważa, powyższe rozumienie wrodzoności zakłada, że istnieje jasno wyznaczona granica pomiędzy przedmiotem biologii i psychologii. Choć naukowcy zakładają, że taka granica istnieje, to jednak nie jest to oczywiste, zwłaszcza kiedy bierzemy pod uwagę obszar poznawczej psychologii rozwojowej – dziedziny „na styku" tych dwóch obszarów. Co więcej, coraz częściej pojawiają się głosy, że nie można kreślić takiej granicy i że poznanie jest inherentnym aspektem życia (Di Paolo, Cuffari, Jaegher, 2018; Kirchhoff, Froese, 2017). Jednak stanowiska, które tak twierdzą, to takie, które porzuciły model reprezentacji jako korespondencji – czyli takie, które pozbyły się teoretycznych ograniczeń płynących z tego paradygmatu, co pozwoliło im na eksplorację procesów powodujących wyłanianie się psychiki.

Jak już zostało omówione przy okazji wrodzoności w biologii, rozwój osobniczy, czy to biologiczny, czy psychologiczny, jest procesem interakcji wielu czynników, zarówno wewnętrznych, jak i zewnętrznych wobec organizmu. Postulat wrodzoności nie jako zakodowania w genach, a jedynie przedmiotowej zewnętrzności z perspektywy psychologii niewiele tutaj zmienia – wydaje się, że nie ma powodów, żeby z góry zakładać, że wyłonienie się normatywności pierwszych reprezentacji zachodzi poprzez interakcję czynników, którymi zajmuja się biologia, chemia lub fizyka, a nie psychologia.

Sama biologia rozwojowa wydaje się podzielać taki pogląd. Armin P. Moczek (2014) np. definiuje rozwój w kontekście biologii rozwojowej jako „sumę wszystkich procesów i wchodzących w interakcję komponentów, które są wymagane, aby forma organizmu i jego funkcja, na wszystkich poziomach organizacji biologicznej, zaczęły istnieć" (s. 218). Zauważmy, że opierając się tej definicji, mamy dwa wyjścia: albo trzymamy się przedstawionego tutaj postulatu, że rozwój w biologii dotyczy jedynie „wszystkich poziomów organizacji *biologicznej*", albo rozszerzymy tę definicję o poziomy organizacji psychologicznej. Jeśli wybierzemy to pierwsze, to automatycznie wykluczamy możliwość, żeby biologia odpowiedziała na proces rozwoju pierwszych reprezentacji – reprezentacje mentalne stanowią element organizacji psychologicznej, nie biologicznej, i stąd ta definicja ich nie obejmie – wyjaśnienie emergencji reprezentacji mentalnych stanie się wtedy obszarem niczyim, odrzucanym zarówno przez psychologię, jak i biologię.

Jednak jeśli rozszerzymy tę definicję i uwzględnimy wyjaśnienie organizacji psychologicznej, nie będzie podstaw – oprócz ograniczeń teoretycznych fundacjonizmu – żeby zakładać, że z wchodzących w interakcje komponentów potrzebnych do wyłonienia się treści mentalnej żaden nie mógłby być studiowany przez psychologa. Scott F. Gilbert (2001; 2015) podkreśla rolę czynników środowiskowych w sumie tych procesów i zwraca uwagę, że rozwój układu nerwowego jest obszarem rozwoju w największym stopniu ze wszystkich aspektów rozwoju osobniczego zależnym od doświadczenia w środowisku. System nerwowy jest ogromnie plastyczny i otwarty na dia-

metralne zmiany na początku życia człowieka i stąd biologicznie „zaprogramowane" schematy aktywacji stanowiące pierwsze reprezentacje są tezą trudną do obrony (Elman, 1996). Mikromorfologia, włączając połączenia neuronalne, jest nastrojona na wpływ środowiska (Pezzulo, Vosgerau, Frith, Hamilton, Heyes, Iriki, Jörntell, König, Nagel, Oudeyer, Rupert, Tramacere, 2015), co ma również sens z adaptacyjnego punktu widzenia, gdyż umożliwia dziecku większą sprawność w nowym środowisku, nawet jeśli jest ono odmienne od środowiska przeszłych pokoleń. W rzeczy samej, jeśli tylko uznamy możliwość, że rozwój pierwszych zjawisk psychicznych zachodzi przy udziale doświadczenia, to nie sposób zaprzeczyć, że ten udział stanowi naturalny obszar badań w psychologii rozwojowej.

Wątpliwe zatem, czy niezbędna dla zjawisk psychicznych normatywność reprezentacji wyłania się tylko i wyłącznie poprzez interakcję czynników z obszarów, którymi zajmują się nauki „niższe" od psychologii. Sprawa oczywiście nie jest przesądzona i choć niepewna, taka możliwość istnieje. Wydaje się, że postulat wrodzoności pojęć nie jest wynikiem rzetelnego podziału obowiązków pomiędzy psychologię i inne nauki, a raczej koniecznym ruchem teoretycznym narzucanym przez ograniczenia inherentne modelowi klasycznej psychologii poznawczej – fundacjonizm.

W kontekście badań ToM jest możliwe, że już niemowlęce zdolności społeczno-poznawcze obserwowane w testach fałszywych przekonań paradygmatu spontanicznej reakcji rozwijają się właśnie w ten sposób, w koniecznym zintegrowaniu czynników środowiskowych – doświadczenia dziecka – oraz czynników wewnętrznych, jego wrodzonej (i.e. „biologicznie" rozwijanej) reaktywności organizmu. Wyniki otrzymane przez Marka Meristę, Karin Strid i Erland Hjelmquist (2016) potwierdzają tę hipotezę (zob. również Meristo i in., 2012). W badaniach tych niesłyszące dzieci nie rozwiązywały FBT spontanicznej reakcji, podczas gdy słyszące dzieci w tym samym wieku tak. Nasuwa się więc wniosek, że nawet ta wczesna kompetencja do rozwiązywania tej wersji FBT nie jest problemem biologa, lecz wyłania się poprzez interakcję dziecka ze środowiskiem, co psychologia rozwojowa może badać. Interpretacja zaproponowana przez Meristę i in. (2016) zwraca uwagę na konieczność interakcji dziecka w obrębie rodziny, aby rozwinęło ono tę podstawową kompetencję.

Wyniki nie są jednak jednoznaczne. Inni badacze (np. Barrett, Broesch,Scott, He, Baillargeon, Di Wu, Bolz, Henrich, Setoh,Wang, Laurence, 2013) przedstawiają dane świadczące o uniwersalności rozwiązywania FBT w paradygmacie spontanicznej reakcji. Oczywiście potrzeba więcej danych empirycznych, żeby wyciągać mocniejsze wnioski. Trzeba jednak pamiętać, że nawet jeśli wszystkie 15-miesięczne dzieci będą rozwiązywać FBT spontanicznej reakcji, to nie będzie to oznaczać prawdy natywizmu. Biorąc pod uwagę interakcyjną naturę zjawisk rozwojowych, będzie to raczej konsekwencją podobieństwa kontekstów rozwoju dla wszystkich dzieci w tym właśnie wieku. Samo stwierdzenie uniwersalności będzie wymagało dalszych badań, żeby ustalić, jak dokładnie zdolność rozwiązywania tego testu się rozwija i czym tak naprawdę jest. Trzeba będzie zbadać naturę poszczególnych kontekstów rozwojowych dzieci i starać się ustalić potencjalne czynniki, które grają rolę w ustalonym uniwersalizmie. Centralne będą tu badania obserwacyjne, skupiające się na pojedynczych dzieciach i sekwencjach ich rozwoju. Dopiero kiedy tego typu badania wykluczą jakąkolwiek rolę czynników środowiskowych w tym zadaniu, będziemy mogli postulować uznanie wrodzoności tej zdolności (i.e. jej emergencję wyjaśnianą całkowicie z perspektywy biologicznej).

PODSUMOWANIE

Klasyczne teorie rozwoju ToM – modularny natywizm, teoria teorii oraz teoria dwóch systemów – operują modelem reprezentacji mentalnej zakładającej, że treść semantyczna jest zakodowana w jej fizycznym nośniku. Ten teoretyczny postulat wymusza pogląd, że reprezentacje nie mogą powstać z niereprezentacyjnych procesów, co prowadzi do konieczności postulowania początkowych reprezentacji stanowiących punkt

wyjścia rozwoju – fundacjonizm – a sam rozwój jest stanowiony jedynie przez „rekonfigurowanie" tych wrodzonych treści. W rezultacie wszystkie powyższe stanowiska są zmuszone do postulowania wrodzonych reprezentacji, choć tylko podejście modularno-natywistyczne mówi o tym otwarcie.

Jedno znaczenie wrodzoności to zapożyczone z biologii „zakodowanie w genach"; pojęcie, które już od jakiegoś czasu w biologii nie funkcjonuje, ze względu na złożoną, wieloelementową naturę zjawisk rozwojowych, której geny są jedynie częścią, i nie jest wobec tego przydatne w psychologii. Drugie, częściej proponowane znaczenie pojęcia wrodzoności w psychologii i kognitywistyce to metodologiczny podział obowiązków – wrodzone ma być to, co wychodzi poza obszar badań psychologii. Jednakże wydaje się, że nie ma powodu, aby sądzić *a priori*, że proces emergencji pierwszych reprezentacji mentalnych (pierwszej normatywności) nie jest stanowiony po części przez procesy, które psychologia może badać. Staje się więc jasne, że głównym motywem uznania postulatu wrodzonych pojęć nie jest żadne z powyższych ustaleń teoretycznych czy metodologicznych, a jedynie ograniczenia przyjętych ram teoretycznych, które nie są w stanie modelować emergencji treści mentalnej ze zjawisk niereprezentacyjnych i są zmuszone do postulowania pierwotnych pojęć, mogących służyć jako źródło inherentnej psychice normatywności.

Obecne rozważania wskazują na potrzebę zmian w teoretycznych podstawach badań nad teorią umysłu. Trzeba wypracować model, który nie będzie wykluczał zjawiska emergencji treści mentalnej podczas ontogenezy *a priori*, i przy pomocy którego będzie można otrzymać pojęcie wrodzoności zgodne z naturą studiowanego przedmiotu – rozwoju ludzkiej psychiki. Podejścia modelujące rozwój poznawczy oparte na działaniu podmiotu (*action-based*, *action-oriented*) stanowią jedną z obiecujących alternatyw (Pezzulo i in. 2015; Allen, Bickhard, 2013), lecz dyskusja ich założeń leży poza zakresem niniejszego tekstu.

PRZYPISY

[1] Test fałszywych przekonań ma na celu ustalenie, czy dziecko rozumie, że osoba posiadająca fałszywe przekonanie będzie zachowywać się według tego przekonania, a nie według znanej dziecku prawdzie (zob. Wimmer, Perner, 1983). Alternatywna wersja testu przeprowadzona w paradygmacie spontanicznej reakcji zastępuje eksplicytne pytanie pomiarem czasu skupienia wzroku podczas każdego z możliwych scenariuszy (zachowanie zgodne z fałszywym przekonaniem i zachowanie zgodne z rzeczywistością). Dłuższy czas skupienia wzroku interpretuje się jako efekt zaskoczenia dziecka, a co za tym idzie jego przeciwnych do obserwowanego scenariusza oczekiwań (zob. Onishi, Baillargeon, 2005).

[2] Wszystkie tłumaczenia cytatów zostały dokonane przez autora tekstu.

BIBLIOGRAFIA

Allen J.W.P., Bickhard M.H. (2013), Stepping off the pendulum: Why only an action-based approach can transcend the nativist–empiricist debate. *Cognitive Development*, *28*(2), 96–133. DOI: https://doi.org/10.1016/j.cogdev.2013.01.002.

Apperly I.A. (2012), *Mindreaders: The Cognitive Basis of 'Theory of Mind'*. Hove: Psychology.

Apperly I.A., Butterfill S.A. (2009), Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116(4), 953–970. DOI: https://doi.org/10.1037/a0016923.

Barrett H.C., Broesch T., Scott R.M., He Z., Baillargeon R., Di Wu, Laurence S. (2013), Early false-belief understanding in traditional non-Western societies. *Proceedings. Biological Sciences*, 280(1755), 20122654. DOI: https://doi.org/10.1098/rspb.2012.2654.

Bickhard M.H. (2001), Why children don't have to solve the frame problems: cognitive representations are not encodings. *Developmental Review*, 21(2), 224–262. DOI: https://doi.org/10.1006/drev.2000.0521.

Bickhard M.H. (2015), What could cognition be if not computation… Or connectionism, or dynamic systems? *Journal of Theoretical and Philosophical Psychology*, 35(1), 53–66. DOI: https://doi.org/10.1037/a0038059.

Bickhard M.H. (2016), Probabilities over what? *Human* Development, 59(1), 34–36. DOI: https://doi.org/10.1159/000447026.

Bickhard M.H., Richie D.M. (1983), *On the Nature of Representation: A Case Study of James Gibson's Theory of Perception / Mark H. Bickhard, D. Michael Richie with Robert Hughes, James Dannemiller*. New York, NY, USA: Praeger.

Bickhard M.H., Terveen L. (1995), *Foundational Issues in Artificial Intelligence and Cognitive Science: Impasse and Solution*. *Advances in Psychology: nr. 109*. Amsterdam, New York: Elsevier.

Butterfill S.A., Apperly I.A. (2013), How to construct a minimal theory of mind. *Mind & Langua*ge, 28(5), 606–637. DOI: https://doi.org/10.1111/mila.12036.

Campbell R.L., Bickhard M.H. (1986), *Knowing Levels and Developmental Stages*. *Contributions to Human Development: nr. 16*. Basel: S. Karger.

Carpendale J.I.M., Wereha T.J. (2013), understanding common developmental timetables across cultures from a developmental systems perspective. *Human Development*, *56*(3), 207–212. DOI: https://doi.org/10.1159/000351276.

Carpendale J.I.M., Hammond S.I., Atwood S. (2013), A Relational developmental systems approach to moral development. W: R.M. Lerner, J.B. Benson (eds.), *Advances in Child Development and Behavior. Embodiment and Epigenesis: Theoretical and Methodological Issues in Understanding the Role of Biology within the Relational Developmental System – Part B: Ontogenetic Dimensions*, nr 45, 125–153. Elsevier. DOI: https://doi.org/10.1016/B978-0-12-397946-9.00006-3.

Carruthers P. (2013), Mindreading in infancy. *Mind & Langua*ge, 28(2), 141–172. DOI: https://doi.org/10.1111/mila.12014.

Carruthers P. (2015), Mindreading in adults: Evaluating two-systems views. *Synthese.* Artykuł w czasopiśmie internetowym. DOI: https://doi.org/10.1007/s11229-015-0792-3.

Di Paolo E.A., Cuffari E.C., Jaegher, H. de (2018), *Linguistic Bodies: The Continuity between Life and Language*. Cambridge, MA: The MIT Press.

Dixson H.G.W., Komugabe-Dixson A.F., Dixson B.J., Low J. (2017), Scaling theory of mind in a small-scale society: a case study from Vanuatu. *Child Development.* Artykuł w czasopiśmie internetowym. DOI: https://doi.org/10.1111/cdev.12919.

Elman J.L. (ed.) (1996), *Neural Network Modeling and Connectionism*. *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: The MIT Press.

Fodor J.A. (1975), *The Language of Thought*. *The Language and Thought Series*. New York: Crowell.

Fodor J.A. (1992), A theory of the child's theory of mind. *Cognition*, 44(3), 283–296. DOI: https://doi.org/10.1016/0010-0277(92)90004-2.

Gilbert S.F. (2001), Ecological developmental biology: developmental biology meets the real world. *Developmental Biology*, 233(1), 1–12. DOI: https://doi.org/10.1006/dbio.2001.0210.

Gilbert S.F., Epel D. (2015), *Ecological Developmental Biology: The Environmental Regulation of Development, Health, and Evolution* (Second edition). Sunderland Massachusetts U.S.A.: Sinauer Associates Inc. Publishers.

Gopnik A. (2003). The theory theory as an alternative to the innateness hypothesis. W: L.M. Antony, N. Hornstein (eds.), *Philosophers and Their Critics. Chomsky and His critics*, 238–254. Malden, MA: Blackwell Pub. DOI: https://doi.org/10.1002/9780470690024.ch10.

Gopnik A. (2009), Rational constructivism: A new way to bridge rationalism and empiricism. *Behavioral and Brain Sciences*, 32(02), 208. DOI: https://doi.org/10.1017/S0140525X0900096X.

Gopnik A. (2010), *Dziecko filozofem. Co dziecięce umysły mówią nam o prawdzie, miłości oraz sensie życia*. *Na ścieżkach umysłu*. Warszawa: Prószyński Media.

Gopnik A. (2011), The Theory Theory 2.0: Probabilistic models and cognitive development. *Child Development Perspectives*, 5(3), 161–163. DOI: https://doi.org/10.1111/j.1750-8606.2011.00179.x.

Gopnik A., Bonawitz E. (2015), Bayesian models of child development. *Wiley Interdisciplinary Reviews. Cognitive Science*, *6*(2), 75–86. DOI: https://doi.org/10.1002/wcs.1330.

Gopnik A., Tenenbaum J.B. (2007), Bayesian networks, Bayesian learning and cognitive development. *Developmental Science*, 10(3), 281–287. DOI: https://doi.org/10.1111/j.1467-7687.2007.00584.x.

Gopnik A., Wellman H.M. (1992), Why the child's theory of mind really is a theory. *Mind & Language*, 7(1–2), 145–171. DOI: https://doi.org/10.1111/j.1468-0017.1992.tb00202.x.

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

Gopnik A., Meltzoff A.N., Kuhl P.K. (1999), *The Scientist in the Crib: What Early Learning Tells us about the Mind*. Pymble, NSW, New York: HarperCollins e-books.

Gould S.J., Lewontin R.C. (1979), The spandrels of san marco and the panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society B: Biological Sciences*, 205(1161), 581–598. DOI: https://doi.org/10.1098/rspb.1979.0086.

Gould S.J., Vrba E.S. (1982), Exaptation – a missing term in the science of form. *Paleobiology*, *8*(01), 4–15. DOI: https://doi.org/10.1017/S0094837300004310.

Helming K.A., Strickland B., Jacob P. (2016), Solving the puzzle about early belief-ascription. *Mind & Language*, *31*(4), 438–469. DOI: https://doi.org/10.1111/mila.12114.

Kirchhoff M.D., Froese T. (2017), Where there is life there is mind: in support of a strong life-mind continuity thesis. *Entropy*, *19*(4), 169. DOI: https://doi.org/10.3390/e19040169.

Kovács Á.M., Téglás E., Endress A.D. (2010), The social sense: Susceptibility to others' beliefs in human infants and adults. *Science (New York, N.Y.)*, *330*(6012), 1830–1834. DOI: https://doi.org/10.1126/science.1190792.

Kristen S., Sodian B. (2014), Theory of mind (tom) in early education: developmental progression of early theory of mind skills, social developmental factors and the importance of tom for learning. W: O.N. Saracho (ed.), *Contemporary Perspectives in Early Childhood Education. Contemporary Perspectives on Research in Theory of Mind in Early Childhood Education*, 291–320. Charlotte, North Carolina: Information Age Publishing.

Leslie A.M., Friedman O., German T.P. (2004), Core mechanisms in "theory of mind". *Trends in Cognitive Sciences*, 8(12), 528–533. DOI: https://doi.org/10.1016/j.tics.2004.10.001.

Lewis C., Carpendale J.I.M., Stack J. (2013), Anticipation and social interaction: Commentary on "Stepping off the pendulum: Why only an action-based approach can transcend the nativist–empiricist debate" by J. Allen and M. Bickhard. *Cognitive Development*, 28(2), 159–163. DOI: https://doi.org/10.1016/j.cogdev.2013.01.008.

Lewkowicz D.J. (2011), The biological implausibility of the nature-nurture dichotomy & what it means for the study of infancy. *Infancy: The Official Journal of the International Society on Infant Studies*, 16(4), 331–367. DOI: https://doi.org/10.1111/j.1532-7078.2011.00079.x.

Lewontin R.C. (2001), *The Triple Helix: Gene, Organism and Environment*. Cambridge (Massachusetts), London: Harvard University Press.

Lightfoot D. (1989), The child's trigger experience: Degree-0 learnability. *Behavioral and Brain Sciences*, 12(02), 321. DOI: https://doi.org/10.1017/S0140525X00048883.

Low J., Apperly I.A., Butterfill S.A., Rakoczy H. (2016), Cognitive architecture of belief reasoning in children and adults: a primer on the two-systems account. *Child Development Perspectives*, 10(3), 184–189. DOI: https://doi.org/10.1111/cdep.12183.

Mameli M., Bateson P. (2011), An evaluation of the concept of innateness. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 366(1563), 436–443. DOI: https://doi.org/10.1098/rstb.2010.0174.

Mayer A., Träuble B. (2012), Synchrony in the onset of mental state understanding across cultures?: A study among children in Samoa. *International Journal of Behavioral Development*, 37(1), 21–28. DOI: https://doi.org/10.1177/0165025412454030.

Meristo M., Morgan G., Geraci A., Iozzi L., Hjelmquist E., Surian L., Siegal M. (2012), Belief attribution in deaf and hearing infants. *Developmental Science*, 15(5), 633–640. DOI: https://doi.org/10.1111/j.1467-7687.2012.01155.x.

Meristo M., Strid K., Hjelmquist E. (2016), Early conversational environment enables spontaneous belief attribution in deaf children. *Cognition*, 157, 139–145. DOI: https://doi.org/10.1016/j.cognition.2016.08.023.

Milligan K., Astington J.W., La Dack (2007), Language and theory of mind: meta-analysis of the relation between language ability and false-belief understanding. *Child Development*, 78(2), 622–646. DOI: https://doi.org/10.1111/j.1467-8624.2007.01018.x.

Mills S. (2001), The idea of different folk psychologies. *International Journal of Philosophical Studies*, 9(4), 501–519. DOI: https://doi.org/10.1080/09672550110081285.

Moczek A. (2014), Towards a theory of development through a theory of developmental evolution. W: A. Minelli, T. Pradeu (eds.), *Towards a Theory of Development*, 218–226. Oxford: Oxford University Press.

Nelson K. (2005), Language pathways into the community of minds. W: J.W. Astington, J.A. Baird (eds.), *Why Language Matters for Theory of Mind*, 26–49. New York, Oxford: Oxford University Press.

Nęcka E., Orzechowski J., Szymura B. (2006), *Psychologia poznawcza*. Warszawa: Wydawnictwo Naukowe PWN; Academica Wydawnictwo SWPS.

Onishi K.H., Baillargeon R. (2005), Do 15-month-old infants understand false beliefs? *Science (New York, N.Y.)*, 308(5719), 255–258. DOI: https://doi.org/10.1126/science.1107621.

Oyama S. (1985/2000), *The Ontogeny of Information*. Duke University Press.

Perner J., Ruffman T. (2005), Psychology. Infants' insight into the mind: how deep? *Science (New York, N.Y.)*, 308(5719), 214–216. DOI: https://doi.org/10.1126/science.1111656.

Pezzulo G., Vosgerau G., Frith U., Hamilton A.F. d. C., Heyes C., Iriki A., Tramacere A. (2015), Acting up: an approach to the study of cognitive development. W: A.K. Engel, K.J. Friston, D. Kragic (eds.), *Strüngmann Forum Reports. The Pragmatic Turn. Toward Action-oriented Views in Cognitive Science* 49–77. Cambridge, MA: The MIT Press.

Piattelli-Palmarini M. (ed.). (1980), *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*. Cambridge, MA: Harvard University Press.

Pigliucci M., Müller G.B. (ed.). (2010), *Evolution – the Extended Synthesis*. Cambridge, MA: The MIT Press.

Pinker S. (2014/1994), *The Language Instinct: How the Mind Creates Language*. Brilliance Audio.

Poulin-Dubois D., Chow V. (2009), The effect of a looker's past reliability on infants' reasoning about beliefs. *Developmental Psychology*, 45(6), 1576–1582. DOI: https://doi.org/10.1037/a0016715.

Racine T.P. (2013), How useful are the concepts "innate" and "adaptation" for explaining human development. *Human Development*, 56(3), 141–146. DOI: https://doi.org/10.1159/000350925.

Ruffman T., Taumoepeau M. (2014), When and how does a theory of mind arise? W: O.N. Saracho (ed.), *Contemporary Perspectives in Early Childhood Education. Contemporary Perspectives on Research in Theory of Mind in Early Childhood Education* 45–63. Charlotte, NC: Information Age Publishing.

Ruffman T., Slade L., Rowlandson K., Rumsey C., Garnham A. (2003), How language relates to belief, desire, and emotion understanding. *Cognitive Development*, 18(2), 139–158. DOI: https://doi.org/10.1016/S0885-2014(03)00002-9.

Samuels R. (2002). Nativism in cognitive science. *Mind & Language*, 17(3), 233–265. DOI: https://doi.org/10.1111/1468-0017.00197.

Samuels R. (2004). Innateness in cognitive science. *Trends in Cognitive Sciences*, *8*(3), 136–141. DOI: https://doi.org/10.1016/j.tics.2004.01.010.

Scholl B.J., Leslie A.M. (2001), Minds, modules, and meta-analysis. *Child Development*, *72*(3), 696–701. DOI: https://doi.org/10.1111/1467-8624.00308.

Scott R.M., Baillargeon R. (2009), Which penguin is this? Attributing false beliefs about object identity at 18 months. *Child Development*, *80*(4), 1172–1196. DOI: https://doi.org/10.1111/j.1467-8624.2009.01324.x.

Scott R.M., Baillargeon R. (2017), Early false-belief understanding. *Trends in Cognitive* Sciences, 21(4), 237–249. DOI: https://doi.org/10.1016/j.tics.2017.01.012.

Scott R.M., Baillargeon R., Song H.-j., Leslie A.M. (2010), Attributing false beliefs about non-obvious properties at 18 months. *Cognitive Psychology*, 61(4), 366–395. DOI: https://doi.org/10.1016/j.cogpsych.2010.09.001.

Spencer J.P., Blumberg M.S., McMurray B., Robinson S.R., Samuelson L.K., Tomblin J.B. (2009), Short arms and talking eggs: Why we should no longer abide the nativist-empiricist debate. *Child Development Perspectives*, 3(2), 79–87. DOI: https://doi.org/10.1111/j.1750-8606.2009.00081.x.

Strijbos D.W., De Bruin L.C. (2013), Universal belief-desire psychology?: A dilemma for theory theory and simulation theory. *Philosophical Psychology*, 26(5), 744–764. DOI: https://doi.org/10.1080/09515089.2012.711034.

Träuble B., Marinović V., Pauen S. (2010), Early theory of mind competencies: do infants understand others' beliefs? *Infancy*, 15(4), 434–444. DOI: https://doi.org/10.1111/j.1532-7078.2009.00025.x.

Venter C., Cohen D. (2004), The century of biology. *New Perspectives Quarterly*, 21(4), 73–77. DOI: https://doi.org/10.1111/j.1540-5842.2004.00701.x.

Villiers J.G. de, de Villiers P.A. (2014), The role of language in theory of mind development. *Topics in Language Disorders*, 34(4), 313–328. DOI: https://doi.org/10.1097/TLD.0000000000000037.
Vinden P.G. (1996), Junin Quechua children's understanding of mind. *Child Development*, 67(4), 1707–1716. DOI: https://doi.org/10.1111/j.1467-8624.1996.tb01822.x.
Wellman H.M. (2014), *Making Minds: How Theory of Mind Develops*. *Oxford Series in Cognitive Development*. Oxford, New York: Oxford University Press.
Wellman H.M., Fang F., Liu D., Zhu L., Liu G. (2006), Scaling of theory-of-mind understandings in Chinese children. *Psychological Science*, 17(12), 1075–1081. DOI: https://doi.org/10.1111/j.1467-9280.2006.01830.x.
Wellman H.M., Fang F., Peterson C.C. (2011), Sequential progressions in a theory-of-mind scale: Longitudinal perspectives. *Child Development*, 82(3), 780–792. DOI: https://doi.org/10.1111/j.1467-8624.2011.01583.x.
Westra E. (2017), Pragmatic development and the false belief task. *Review of Philosophy and Psychology*, 8(2), 235–257. DOI: https://doi.org/10.1007/s13164-016-0320-5.
Westra E., Carruthers P. (2017), Pragmatic development explains the Theory-of-Mind Scale. *Cognition*, 158, 165–176. DOI: https://doi.org/10.1016/j.cognition.2016.10.021.
Wimmer H., Perner J. (1983), Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103–128. DOI: https://doi.org/10.1016/0010-0277(83)90004-5.
Wynn K. (1992), Addition and subtraction by human infants. *Nature*, *358*(6389), 749–750. DOI: https://doi.org/10.1038/358749a0.

Nota autorska
Niniejszy tekst został przygotowany w ramach projektu badawczego „O społeczno-kulturowym uwikłaniu potocznych teorii umysłu. Dyskusja o granicach uniwersalności i powszechności pojęć i atrybucji mentalnych", finansowanego przez Narodowe Centrum Nauki (grant PRELUDIUM 12, UMO-2016/23/N/HS1/02887 na lata 2017–-2019).

Pragnę podziękować Arkadiuszowi Gutowi, Mateuszowi Chrólowi oraz dwóm anonimowym recenzentom za trafne uwagi, które pozwoliły na ulepszenie pierwotnej wersji tekstu.

frontiers | Frontiers in Developmental Psychology

TYPE Conceptual Analysis
PUBLISHED 15 August 2024
DOI 10.3389/fdpys.2024.1449705

OPEN ACCESS

EDITED BY
Stephanie M. Carlson,
University of Minnesota Twin Cities,
United States

REVIEWED BY
Michael Tomasello,
Duke University, United States
Philip D. Zelazo,
University of Minnesota Twin Cities,
United States

*CORRESPONDENCE
Jedediah W. P. Allen
jallen@bilkent.edu.tr

RECEIVED 15 June 2024
ACCEPTED 29 July 2024
PUBLISHED 15 August 2024

CITATION
Allen JWP, Mirski R and Bickhard MH (2024) Beyond the mirror: an action-based model of knowing through reflection.
Front. Dev. Psychol. 2:1449705.
doi: 10.3389/fdpys.2024.1449705

Beyond the mirror: an action-based model of knowing through reflection

Jedediah W. P. Allen[1]*, Robert Mirski[2] and Mark H. Bickhard[3]

[1]Bil-Ge Lab, Department of Psychology, Bilkent University, Ankara, Türkiye, [2]Department of Cognitive Science, Nicolaus Copernicus University in Toruń, Toruń, Poland, [3]Department of Philosophy and Psychology, Lehigh University, Bethlehem, PA, United States

Epistemic reflection involves the creation of qualitatively new knowledge. Different models have been proposed to account for new knowing through reflection that have typically been grounded in an information-processing framework. However, there are in-principle arguments that information-processing approaches preclude the emergence of new representation altogether. Accordingly, any information-processing account of knowing through reflection is plagued by emergence issues. After discussing some of these emergence issues for four prominent models in the cognitive science literature, an alternative action-based model of representation and reflection is presented called interactivism. Interactivism's model of representation, as grounded in action anticipations, serves as the foundational emergence needed to account for subsequent knowing through reflection. After introducing the interactivist models of representation and reflection through knowing levels, some of the implications for consciousness, enculturation, language, and developmental methodology are discussed.

KEYWORDS

epistemic reflection, action-based approach, interactivism, emergent representation, interactive knowing, enculturation, language as interaction

1 Introduction

Reflection is often characterized as serving one of two functions: the creation of qualitatively new knowledge, or qualitatively new capabilities involving self-/emotion-regulation through some sort of distancing process. While most researchers incorporate some role for language in the reflection process, a basic division can also be drawn between approaches that emphasize the developmental origins of reflection as a cognitive activity vs. those who argue that language is the original locus through which reflection takes place. In the current paper, we will explore efforts to explain the development of reflection as a cognitive activity for emergent knowing, but we will also indicate the subsequent role that language plays in this process. The paper will proceed by briefly discussing several different models that are all united in trying to explain how reflection enables the creation of qualitatively new knowing: these include Mandler's (2004) perceptual analysis, Karmiloff-Smith's (1992) representational redescription, Perner's (1991) meta-representation, and finally, Zelazo's (2004) levels of consciousness model. This discussion will ultimately reject the adequacy of these models due to their information-processing assumptions and inability to account for representational emergence. The alternative interactivist model is an action-based framework that contrasts with an information-processing

ontology (Bickhard and Terveen, 1995; Bickhard, 2024)[1]. This model will be introduced and discussed in the context of interactive vs. reflective knowing, primary consciousness vs. reflective consciousness, and internalization vs. enculturation as the process of socialization. Finally, the role of language for reflection will be addressed in terms of its differential relevance for both pre-reflective and reflective development with some implications for developmental methodology.

2 Qualitatively new knowing: existing models (Mandler, Karmiloff-Smith, Perner, and Zelazo)

At the core of the developmental sciences are issues of origins in general and the issue of representational/knowing origins in particular. Nativist positions generally side-step the issue of origins by assuming that essential knowledge structures are provided to the species through some unspecified evolutionary process. The theoretical motivation for nativism comes from learnability arguments that innateness is *necessary* for learning to get started (Chomsky, 1959; Fodor, 1975). Contemporary empiricist positions actually agree with the need for some innateness but disagree about the amount and type (e.g., feature representations or full concepts, a few concepts or many concepts, for a full discussion see Allen and Bickhard, 2013). However, empiricist positions are more developmental and pursue more powerful possibilities for learning such that they assume that qualitatively new representations/knowing are produced during ontogeny. That said, both nativist and empiricist positions tend to assume a background information-processing framework with implications for the nature of representation that make qualitatively new representations (i.e., emergence) impossible. For information-processing approaches, the nature of representation is in terms of some sort of encoding relationship with the world (Bickhard and Terveen, 1995; Bickhard, 2009a).

Encodingism is the assumption that foundational representations are encodings. Encodings are constituted by a correspondence relationship with what they represent, and these correspondences are often assumed to be causal, nomological, or informational. Regardless of the specific nature of the relationship, encodings are representational stand-ins such that they must derive their content from some other source of representation. For example, the rings in a tree encode its age in years. This is a factual/informational[2] relationship that is only representational if there is an epistemic agent who already knows about rings in trees, annual growth, and the relationship between the two. Without an interpreting agent, there is no content for the encoding and its relationship to what it represents. Thus, as an account of foundational representations, encoding approaches are incoherent. What's needed is an account of emergent representation in which representation is emergent within a foundation that is not already representational and only action-based approaches have offered to provide such account (Allen and Bickhard, 2013). While Piaget is the best known action-based approach, it is the interactivist model that will be presented in Section 2. Before that discussion, we present four empiricist models that all assume that learning and development involve qualitative changes in the nature of the representations that can be constructed through reflection; however, all four models are also committed to an information-processing framework that precludes the possibility of emergent representation.

2.1 Perceptual analysis (Mandler)

Jean Mandler's model has the laudable goal of trying to account for the foundations of meaning itself (Mandler, 2000, 2004). We refer to this issue as the *foundational emergence* problem. One avenue for resolving this issue is to take a nativist stance (i.e., foundational meanings/representations are innate). However, such an approach does not solve the issue so much as it ignores it. This leaves empiricism as the alternative—and in the current state of the field, some form of information-processing empiricism. While modern empiricist approaches to development also start with some amount of innate conceptual/representational base (Gopnik, 2003), Mandler's model attempts to address the developmental emergence of such a base. This model commits to a sharp distinction between perception and (meaningful) conception, in which the latter is supposed to derive from the former through an abstraction process—*perceptual analysis*. Perceptual analysis is a volitional process involving attention to certain aspects of the perceptual data stream in order to abstract a simplified rendering of the input. This process also involves "recoding" the format of the abstracted content into "explicit" form which enables "…one to describe, recall, or think about something new, not just recognize it" (Mandler, 2004, p. 18). Mandler grounds much of her theorizing in a set of empirical findings in which very young infants seem to have abstract (i.e., conceptual) categorization abilities that include the functions of objects (in addition to their perceptual features).

While these empirical findings should be a constraint on any adequate theory of foundational meaning, Mandler's account has a number of theoretical problems that appear to be unsolvable (see Müller and Overton, 1998 for a full treatment of the model and its limitations). Two of the most relevant of these problems concern the abstraction process. Abstraction is supposed to produce meaningful/conceptual categories. But how can the correct features be abstracted without already knowing what the

1 Tomasello (2024) has recently proposed an agency-based model of reflection that is also more pragmatically oriented in its background assumptions. Interestingly, it also shares the basic distinction between what we call interactive vs. reflective knowing in terms of executive vs. metacognitive regulation.

2 We contrast the meaning of "information" in information theory with the semantic meaning of it—as mental representation with truth value, intentionality, and content. Informational relationships in the former sense are integrated into the interactivist model in terms of epistemic contact (i.e., differentiation/detection) not epistemic content (i.e., representation, Bickhard, 2009a). Detected correlations need to be accompanied by anticipations, otherwise they mean nothing to the organism. In this paper, when we refer to "information," it is in the information theoretic sense, unless stated otherwise.

category is supposed to be? Mandler's empirical works suggests that perceptual features like object salience or similarity are not sufficient—correct abstraction requires already knowing which features should be abstracted. For the second issue, abstracting relevant features means taking a subset of the perceptual data, but a subset of the input stream does not give new data. A subset may give new access to volitional processes, but there is no new data *per se*. Further, both problems assume that feature representations are available as distinct "pieces" of an overall representation (i.e., a feature nativism). However, perceptual analysis was presented as an alternative that could avoid the pitfalls of nativism. This means that in addition to the problems with abstraction as an account of new content, the need for a feature innateness/foundationalism means that such a model is open to the same emergence limitation as standard nativist accounts—that representations *must* already be assumed in order to explain the origins of new representations, whether in development or in evolution (Allen and Bickhard, 2013). Lastly, a third issue is that changing formats does not change the content of the data, nor does it make the data more/less explicit (more on this in the discussion of Karmiloff-Smith). The interactivist account of representation in the current perspective will provide a model for foundational emergence that does not have the above problems.

2.2 Representational redescription (Karmiloff-Smith)

Karmiloff-Smith's model builds on the work by Mandler in terms of foundational meanings to explain a process for the subsequent development of new representations (Karmiloff-Smith, 1992). Karmiloff-Smith accepted much of the cutting-edge empirical conclusions coming out of the nativist research program in the 80 and 90's while also trying to transcend the contrast between domain-*specific* and domain-*general* learning processes (Karmiloff-Smith, 2018). Her domain-*relevant* approach attempted to explain how innate biases could result in a cascade of emergent developmental outcomes. Accordingly, Karmiloff-Smith provides an account for the emergence of new forms of representation that go beyond the foundational emergence of conceptual from perceptual. We refer to this issue as the *subsequent emergence* problem. Similar to Mandler, Karmiloff-Smith also appeals to changes in format to account for new content. Different from Mandler, Karmiloff-Smith adopts a more robust constructivism in that there are *internal sources* of change such that cognitive processes derive new content from the overall organization of old content (e.g., information embedded in problem-solving procedures). This is a process of *Representational Redescription* (RR) in which the implicit content of prior knowing is made explicit and constitutes new representational content.

The main function of RR is to facilitate flexibility, and thus, control of behavior relative to new purposes. This is a consequence of the increased operations that can *access* the more explicit representations that eventuate in conscious access, linguistic access, and theory construction processes (cross-domain integration). The RR process suggests four types of representations: one implicit (I) and three explicit (E1—unconscious, E2—conscious access, and E3—conscious and linguistic access). Implicit representations are procedures (or sensorimotor encodings) that have a sequential organization that is encapsulated and inflexible. These representations are used in response to external stimuli (i.e., they are not internally driven). The RR process involves *reformatting* the sensorimotor encodings through abstraction so that more operations can access their contents. It is an abstraction in the sense of extracting the sensorimotor information while losing the perceptual details. Once the newly formed E1 representations are available, they can be used in more flexible ways (e.g., understanding the analogy between a Zebra and a crosswalk sign). This means that the creation of E1 representations precedes any sort of reflection about potential relations embedded in the sensorimotor procedures. At E2, representations are in a format accessible to consciousness but not verbal report. Finally, E3 representations are needed to use language because they involve a "cross-system code." This makes language a tool that can be used after two or three iterations of the RR process have abstracted them into the correct format.

Karmiloff-Smith's theorizing involves developmental elaboration beyond the model of foundational meaning provided by Mandler's account. This elaboration is both at the broader level of development and at the specific level of RR. Similar to the perceptual analysis account, the issues for abstraction as a source of new content apply here as well. However, the focus of RR is on how changes in format affect explicitness, which enables new forms of knowing. There are two issues here: (1) does format affect the explicitness of the representational content? (2) does format enable new forms of knowing? For (1), as Fodor (1998) indicated, all encodings are explicit about something and so the idea of implicit representation cannot be with respect to the content of the representation itself. For example, changing the format of the letter "S" to "…" does not alter the explicitness of either representation[3]. What changes from "S" to "…" are the sorts of things one can do with the new encoding (Bickhard and Terveen, 1995). The dots can be sent over telegraph lines while the letter cannot. Accordingly, for the RR model, the implicitness is in terms of how the overall systems can (or cannot) make use of the (explicit) content of the "implicit" representations. This means that the changes in format from sensorimotor encodings to E1 do not involve new content for the E1 representation (or E1 to E2 or E3). However, if the changes in format through the RR process do not involve the emergence of new content, then the increasing access does not involve new forms of knowing. That is, issue (2) is also answered in the negative.

At the broader level of development, Karmiloff-Smith has captured several important features. Her theorizing suggests that the internal dynamics of cognition are a source for change with recurrent phases of development that oscillate between behavioral mastery and cognitive reorganization. This makes it important to consider how the same behavioral performance at two different ages may in fact be a consequence of different cognitive processes. This means that *U*-shaped development is not noise to be averaged away but an important constraint on developmental

3 Encoding content is borrowed from or defined in terms of other already available content—e.g., "…" from "S." It cannot create new content. That is the central problem with "information" processing models.

explanations (see also Gershkoff-Stowe and Thelen, 2004). The theory also makes multiple distinctions about different forms of knowing. Representational multiplicity is important because there is a strong tendency in development psychology to ignore the possibility that children at different ages have qualitatively different ways of knowing (adultocentricism) and to thus not control for such possibilities in "empirical" research (Allen and Bickhard, 2013). Finally, Karmiloff-Smith's theory attempts to reconcile the emergent constructivism of Piaget's theory with the representational innateness of nativist research programs. In this respect, it shares an overall goal and structure with Carey's (2009) more recent model of how to reconcile innateness with qualitative development. However, in both cases, the requirement of an innate representational foundation for learning and development involves a notion of representation that precludes the possibility of genuinely new representational content (i.e., encodingism). Further, an adequate account of new content through learning obviates the necessity for an innate foundation. Thus, either qualitative emergence in development is impossible, or, there is no necessity for (homuncular) innateness (Allen and Bickhard, 2011).

2.3 Meta-representation (Perner)

Perner (1991) has developed a model of meta-representational development to account for changes in false-belief understanding and a number of other qualitative changes around age 4. This model suggests that meta-representational development involves new knowledge in that children become able to represent representational relationships, and this has cascading developmental consequences. In particular, children with meta-representational abilities are able to understand misrepresentation (of people with false-beliefs or objects like signs and photographs), the representational nature of language (i.e., that words are not properties of what they represent), and the distinction between sense and reference as manifest in understanding that Clark Kent and Superman are the same person (Perner et al., 2002; Iao et al., 2011). Although this model has some basic convergence with the interactivist model to be presented below, it has been discussed in detail from the interactivist perspective previously (Bickhard, 1992). The most relevant conclusions from that discussion are that no account of foundational emergence will be possible given the (encoding) assumptions about representation and that reflection seems to already be needed for even the basic representations of infants (not just the meta-representing of preschoolers).

2.4 Levels of consciousness (Zelazo)

A more recent model for how the development of reflection enables new forms of knowing, representing, and acting comes from Zelazo (1999, 2004, 2015). This model is similar to Karmiloff-Smith's in that it is: focused on levels of subsequent emergence, developmentally rich, conceptually coherent, and grounded in both behavioral and neural data. It is also unique in terms of the focus on consciousness as being relevant for modeling changes in knowing. Nonetheless, as with Mandler, Karmiloff-Smith, and Perner, the underlying information-processing empiricism creates limitations for how well the model can account for epistemic reflection (i.e., the emergence of new knowing through reflection).

Much of the recent empirical motivation for the "emergence" process in this model comes from brain studies in which there seems to be "iterative reprocessing" of information within and between areas of the brain (Zelazo, 2015). However, if the technical sense of information relevant for brain studies cannot account for the semantic sense of information relevant for cognition, then the implications of these data are unclear. Further, the myriad reciprocal projections of the brain can also be characterized as supporting oscillatory processes, rather than semantic "re-entrant" processes, and oscillatory processes have been argued as a neural foundation for the anticipatory processes that constitute the core of an action-based "semantics" (Bickhard, 2015, 2024).

Regardless of the status of re-entrant processing, the original reflection model is mostly explicated in terms of theoretical considerations, and that will be the focus of our analysis. The Levels of Consciousness (LoC) model is an account of changes in the reflective capabilities of children (Zelazo, 1999, 2004). New reflective capabilities enable more complex representing through the creation of new representations (i.e., of relations between lower-level representations and of hierarchical control structures). Zelazo highlights intentionality as the key feature of any form of consciousness. This is intentionality in Brentano's sense of being *about* something and for motivating action [1973 as cited by Zelazo (2004)], but there is no account of the emergence of intentionality itself. Instead, intentional representations/descriptions of objects in semantic Long-Term Memory (LTM) are triggered by actual objects from the environment. These representations then trigger the most salient action pattern that has been learned through association (e.g., a rattle might trigger the action pattern of sucking at one age and shaking at another). This form of representing is supposed to constitute basic consciousness (i.e., minimal Consciousness or minC).

Although the mechanism for ascent in the LoC model is the same (i.e., recursion), the most qualitative change in representing takes place in the transition from level 1 to level 2 at the end of the 1st year of development. This change involves a constitutive role for language in terms of labeling. Labels are supposed to provide an enduring trace to segments of the perceptual input stream that constitutes basic conscious experience (i.e., minC). These traces are representations proper in that they can be "decoupled" from the ephemeral flow of experience and manipulated in working memory as part of top-down control (e.g., representation of an occluded object that can serve as a goal). However, for labeling to serve this decoupling function requires level 2 consciousness to create identity relations between two moments in the input stream from first-level consciousness. Thus, the construction of these identity relations require reflection through *recursion*. Recursion is understood in the sense of a computer program that calls on itself as a parameter [e.g., Factorial (n) = n * Factorial (n-1)]. More recent discussion about reflection is in terms of *iterative reprocessing* where information output is fed back into the system to be combined and integrated with existing representations to create a new interpretation of the situation (Zelazo, 2015).

Our concern with this model can be divided into two issues: (1) how do semantic representations/descriptions work such that labels liberate the infant from the flow of first-level consciousness? (2) how does recursion enable new levels of consciousness? We suggest that the answers given by the model presuppose a rich innate representational base as well as the reflection capability that was meant to be explained. First, labels (from semantic LTM) are supposed to be attached to identity relations that connect the contents (also from semantic LTM) from two moments in the input stream. However, this process seems to be creating a linguistic encoding of the content of the identity relation with the label—instead of "…" there is "dog" whose content is dog, and the content of dog came from semantic LTM. This means that all of the content for the encoding relationship is coming from semantic LTM with no account of its origins or how the semantic descriptions are being interpreted in the first place. Further, if reflection is needed to make the new linguistic encodings (in addition to it being needed to create the identity relations and perhaps for interpreting the descriptions in the first place), then this leaves recursion to account for all of the functionality of reflection[4].

Second, if reflection is required to both interpret semantic descriptions and attach them to labels (recC) or to objects (minC), then reflection is present from the very beginning, and this would make it homuncular (Bickhard and Terveen, 1995). If reflection were already present, then perhaps recursive/re-entrant processing could construct something "new." That is, if semantic information contents are re-entered into a consciousness that is already reflective, then a homunculus can survey all those contents (with all of the consequences at each level that the model posits). However, this would not create new content, instead, different levels of detail are being selected with different levels of reflection. This makes the development of "new" representation a matter of *selection* amongst existing content rather than the *emergence* of new content[5]. If our analysis is correct, the LoC model is not able to fulfill its epistemic function. This is because recursion does not yield a higher level of consciousness *per se*, but yields a hierarchy of levels of "content" within reflection. This may be the best option available within an information-processing framework but that is not the only alternative for modeling development.

As an account of emergent forms of knowing through reflection, the LoC model appears problematic; however, the descriptions, properties, and functions attributed to the different levels of consciousness may still capture something important about development. That is, the LoC model may be adequate for certain aspects of the developmental changes in consciousness even if it is not adequate as a model of epistemic reflection. Further, a core feature of all of the models reviewed above is the idea that lower-level representations serve as the foundation for new representations at higher levels through reflection. The current interactivist model of reflection shares this basic idea but the crucial difference concerns its action-based foundation (Allen and Bickhard, 2013). In contrast, all of the above models are developed within an information-processing empiricist framework. This framework is incapable of accounting for emergent representation and precludes the possibility of an emergent constructivism (Bickhard and Terveen, 1995; Allen and Bickhard, 2022). Without an emergent constructivism, learning and development cannot result in new knowing, and any model of reflection will ultimately fail as an explanation for such an outcome.

3 Interactive knowing and reflection

Interactivism is an action-based model of cognition and persons in which knowing is doing, and competent knowing means successful interaction (Bickhard, 2009b, 2024). Perhaps the best known action-based approach in developmental psychology was Piaget's sensorimotor theory (Piaget, 1954). However, misinterpretations and misguided methodology side-lined Piagetian theory in general and its action-orientation along with it (Smith, 1993; Allen and Bickhard, 2013). Rejections of computationalism for some strands of cognitive science have seen a move toward embodiment and most recently an explicitly pragmatist turn (Engel et al., 2016). However, interactivism differs from these embodied/pragmatist approaches, including Piaget's, in terms of the underlying models of representation (i.e., interactive knowing) and reflection (Bickhard, 1978; Campbell and Bickhard, 1986; Bickhard and Terveen, 1995).

For interactivism, representation is constituted in terms of anticipating potential interactions with the world. The anticipations are discovered to be true or false once enacted (i.e., they have *truth-value*) and they involve presupposition that the world will cooperate (i.e., they are *about* the world). For example, to anticipate that a coffee cup can be picked up presupposes that the cup is not broken. Being unbroken is usually presupposed by our interactions with cups, but it is not indicated within the anticipation and therefore it is not represented explicitly. However, if that presupposition is relevant (i.e., the cup is in fact broken), then the interaction will fail (or at least break down) and thus, presuppositions can be functionally important for the interaction without being explicitly represented. In this model, presupposition provides the implicit content that is about the world (note that presupposition is an aboutness that is not homuncular) while the explicit content is constituted in the internal anticipations or indications of potential interactions *per se* [e.g., a "pointer" indication of a subsystem that could engage in the anticipated-as-possible interaction(s)][6].

Let us stress the point that interactivist mental content is constituted by what is *implicitly presupposed* by the anticipatory dynamics, which contrasts with the criticized model of

4 There are also potential empirical reasons for caution about the role of labeling in this model as it is not clear that infants use labels to succeed on tasks like A-not-B at the end of the 1st year, or what kind of labels those would be Müller and Kerns (2015); also, non-human animals seem to have rather sophisticated forms of top-down control although they do not use language (Penn et al., 2008).

5 Further, how could reflection explain the origins for how we represent non-observables like mental-states. No amount of reprocessing at any level of resolution is going to enable the extraction of something that is not already present in the input stream of conscious experience.

6 The possibility of pointers show that indications pose no particular problem, although that is not how the CNS actually does it. See Bickhard (2015, 2024) for how the indicating/anticipation function is served in the CNS.

encodingism. As we have discussed earlier, encodingism views mental content as constituted by information in information theoretic sense, i.e., by correlation between the agent's internal states and some feature of the world (see text footnote 2). In the interactivist critique of encodingism, the issue is not whether or not information plays a role in cognition. Information understood as correlation is a property of the world and it naturally matters to agents. Rather, the problem is the ontological assumption that information *constitutes mental content*. One of the critical points we made earlier is especially relevant here—correlation needs to be known in order to be representationally utilized by the agent and so it cannot be what constitutes that knowledge itself. In contrast, content as implicit presupposition makes no such problematic assumption; as a natural consequence of learning to effectively interact with the world, the organism's anticipatory knowledge comes to "agree" with how the world is, to implicitly presuppose how it is.

For a developmental example, consider object representation. Object representation for the 2-year-old is constituted in the web of anticipated possibilities for interaction remaining constant with respect to other sorts of changes (e.g., occlusion or displacement). While the basic properties of representing are present in the anticipations (i.e., truth-value and aboutness), the permanence is a property of the overall organization of the web of anticipated possibilities. Such permanence is functional for the 2-year-old in that they can act in accordance with the presupposition that the object has a continued existence, but the permanence *per se*, the presupposition, is not itself represented by the toddler. This is because the toddler cannot directly interact with the permanence of the object and therefore cannot have anticipations directly about it. Instead, reflection will be the process that enables the implicit content/presupposition to become explicit (i.e., reflection is required to know about permanence *per se*).

Interactive knowing is constituted in the organism/system interacting with the environment (i.e., first level knowing). Reflection requires a second interactive system that can interact with the first system/organism (i.e., second level knowing). In humans, this means that the development of reflection involves an architectural change to the CNS—maturational development of the brain—to enable interaction between regions (i.e., second level knowing) in a fashion similar to how the CNS of the toddler interacts with the world (first level knowing, Bickhard, 2015, 2024)[7]. With reflection comes the possibility of knowing about the system (its internal functional organization) that interactively knows the world. That is, the properties and relations implicit in first level knowing (i.e., the presuppositions of interactive knowing) become knowable through reflection (i.e., second level knowing). While there are no a priori constraints on the age of development for reflection, there are ample empirical reasons to think that it is around age 3.5–4 (Allen and Bickhard, 2018). This is the age at which there seems to be developmental transitions in abilities within and across domains. There is also evidence that uniquely supports the interactivist model of reflection over other domain-general explanations for such changes at this age (see discussion of Allen et al., 2021 below).

To further illustrate the contrast between interactive and reflective knowing, let us consider the development of an empirical test specific to the interactivist model of reflection. Any such test is difficult for three general reasons: first, given that any task can, in principle, be interactively learned through non-reflective knowing, it is important that the task have sufficient novelty. Second, if all the different interactions of a toddler[8] are already consistent with the implicitly presupposed properties like permanence, then what difference does it make to have explicit knowledge of those presupposed properties? Third, as adults, our reflectively conscious experience of objects can always be explicit, and so it can seem as if infant interactions that are consistent with our explicit representations are also explicit for the infant (i.e., adultocentrism).

To address these issues, a test of reflection was developed that turned on being able to explicitly represent the relationship between two objects—a mutual support relationship (Allen and Bickhard, 2018). Similar to the permanence of objects, relations amongst objects cannot be directly interacted with and therefore cannot be explicitly represented by toddlers. Without representing relations *per se*, children should not be able to anticipate their consequences in a sufficiently novel situation. Accordingly, the Leaning Blocks (LB) task involves asking children what will happen to a block being held at a 45° angle when released (i.e., "fall" or "stay up"). After asking the same question for a second block, the test question involves holding the two blocks such that they are *leaning* against each other. Children are again asked what will happen upon release. Three-year-old's fail the question while 4- and 5-year-olds are basically at ceiling. These findings suggest that the older children can explicitly represent the mutual support relationship between the two blocks, and in so doing, correctly determine the consequence given the relative novelty of the situation. A follow-up study, that included a second reflection task (i.e., Candy Monster, CM) and three EF measures, suggests that the results from LB are not due to changes in executive functions. Specifically, inhibition, working memory, and cognitive flexibility interpretations were tested against the reflection interpretation and the results favored the later (Allen et al., 2021). Importantly though, reflection is an enabling constraint which means that learning relevant to any specific task must still take place before the "reflective ability" can be measured. The design intention of the LB and CM tasks are as relatively "pure" measures of reflection because they do not seem to involve many additional abilities beyond explicit representing *per se*.

7 For example, a maturation of a neural loop from pre-frontal to basal ganglia to thalamus and back to cortex (Bickhard, 2015, 2024), thus enabling interaction with other regions of the CNS.

8 It is not until toddlerhood that children show a coherent set of interactions consistent with the permanence of object. At earlier ages, infants show only a limited set of interactions consistent with permanence (Baillargeon, 2008). For example, small changes in whether a looking measure involves occluding an object vs. covering it, and later, containing it, affects performance such that the same aged infants fail one version while passing the other(s).

3.1 Consciousness and reflection

"Consciousness" is often used in a crucially equivocal manner: (1) as an "awareness" of the potentialities that constitute the world, and (2) as a kind of reflection on those first level processes and organizations. Failing to distinguish these yields aporetic problems in understanding consciousness (Bickhard, 2005). For example, as Dewey pointed out about Russell's "sense data" (Dewey, 1915, 1941; Tiles, 1990), sense data (today's descendent is "qualia") are supposed to *constitute* "consciousness" of the world, but in fact sense data (qualia) are products of *analysis* of (reflection on) primary awareness—they are generated in analysis, not constitutive of what is being analyzed,

In the interactivist model, there is a clear distinction between first level interactions with the environment and anticipations of possible such interactions, and second level interactive reflections on those first level processes and properties (and relations). The model of primary awareness has already been outlined: anticipations of (organizations of) possible interactions and their intrinsically related presuppositions. The model of reflection is that of a second level of such interactive "knowing" that interacts with the first level. The first iteration of such reflection is not possible in all species—it requires the macro-evolution of a special functional organization in the brain, and a developmental maturation of that functional organization in the individual. Further levels can be constructed in a strictly functional manner through language and culture (Bickhard, 2024), which will be discussed briefly in what follows.

4 Internalization vs. enculturation

While psychology today generally accepts that human minds are largely shaped by culture, the actual models of how that happens remain problematic (Turner, 1994, 2018; Christopher and Bickhard, 2007). Culture tends to be framed in terms of a set of beliefs and practices that the child "internalizes" as she undergoes the process of enculturation. The concept of internalization can be traced back to both Piaget (1952), Piaget and Inhelder (2000), and Vygotsky (1978), but its current uses usually draw on the latter. Vygotsky was especially interested in internalization of culture. His idea was that culture is dialectically externalized and internalized by any individual interacting socially. Children, being newcomers to social reality, were said to internalize into their minds the ways of thinking instantiated in social interactions, which made for the central mechanism of enculturation in his theory.

The details of the presumed internalization process remain vague; most fundamentally, the question arises as to what it actually means—how something that is out there in the world can get into the child's mind? And once it gets there, what kind of phenomenon is it? Potential answers to these questions depend on one's wider ontology of the mind. In encodingist models, which still dominate the field, the internalization process has been argued to be a conceptually incoherent proposal (Christopher and Bickhard, 2007). This incoherence is a consequence of the wider problems with encodingism discussed earlier: In order to internalize anything that is outside of the agent—such as a norm or custom—an encodingist agent would have to already know the thing in order to be able to internalize it, which means that internalization cannot be the basic mechanism for how cultural knowledge is formed (cf. the similar critique by Piaget, 1971). The interactivist model of enculturation, in contrast, follows naturally from the principles on which the interactivist ontology is based, and has no need for the concept of internalization.

Enculturation in interactivism follows the same basic principles as development of interactive knowledge of the physical reality—what differs is the object of interaction and resulting anticipatory organization: While knowledge of the physical world is constructed by engaging with and learning, for instance, the interactive stabilities of physical objects, cultural knowledge originates in the child's interaction with cultural or conventional objects of social ontology, such as norms governing dinner or nighttime routines (for the interactivist model of social ontology as convention see Bickhard, 1980; Mirski and Bickhard, 2021). Consequently, the pre-reflective knowledge of a child developing within a culture involves implicit presuppositions about cultural phenomena—it is organized in a way that "honors" cultural phenomena such as values or customs, but the child does not represent them explicitly as such; culture is implicit within the child's anticipatory organization, it is part of how the person views the world and interacts with it. Rather than internalization, the process is that of construction constrained and guided by the socio-cultural milieu.

Implicit presuppositions concerning the socio-cultural world, similarly to those concerning physical reality, can be represented explicitly once reflection is available to the child. For example, at knowing level 1, the child can interactively differentiate him or herself from other agents and the rest of the world, but she will not be able to represent that differentiation explicitly. In other words, the child will have a self, but will not know it. This implicit self will be greatly constrained and guided by culture as it will involve all types of presuppositions about the social world and its norms, such as, for instance, a preference to play with toy cars rather than dolls. Reflection, or level-2 knowing, allows the child to examine the self-embodied in level-1 organization and develop, for instance, meta-strategies for navigating the social world, such as heuristics for successfully creating play situations with toy cars rather than dolls. These reflective representations and strategies will constitute the child's self-representation, or its *identity*—a set of ways of being in the world. However, this self-representation will not be known explicitly, the child will not be able to represent the way it represents him or herself—for that, a third level reflection is needed. The self-representation will have their own set of implicit presuppositions, which again can be only explicitly known by a higher level of knowing; once that is available, the child will be able to, for instance, compare her own identity with alternatives or examine it in terms of values and perhaps reconstruct it to agree with them (Campbell and Bickhard, 1986, p. 118–127). The epistemic climb up the knowing levels need not stop at level 3—every epistemic level involves its own implicit presuppositions, which can be potentially known by a level higher than that. A level 4 examination of one's identity may involve a discovery that one has a tendency to frequently switch between identities, which can then be duly addressed by the agent. Importantly, even though every level leads to the emergence of qualitatively new knowledge, it too involves implicit presuppositions that remain unknown before a higher-level reflection makes them explicitly. While there is not

an in-principle limit to how high in the reflective levels the agent can climb, there naturally are various factors that influence the process[9]. Among them, language seems to be a major one, to which we turn below.

5 Does language serve a reflective function?

Interactivism models language as a system for interacting with social situations, or situation conventions, which constitute social reality (Bickhard, 1980). The basic idea is that language is a meta-convention—a convention for interacting with conventions—that allows the agent to coordinate action with its conspecifics. For instance, consider the child's early developing use of the utterance "no!" and how he or she uses it to negotiate or modify social situations—even though at first the child uses it simply to protest the current situation, it is understood by both the child and the caregiver in a similar way and thus succeeds in communicating the desired change to the situation (i.e., that it should stop or change). Importantly, such early uses of language are fully implicit and do not amount to a symbolic understanding of utterances—they are part of knowing level 1[10].

However, pre-reflective mastery of language is limited: language is not in this early form understood symbolically, i.e., as representing some part or aspect of reality, but only as yet another way of interacting with the world. As such, it does have presuppositions about it, and—just like any other knowing in interactivism—those presuppositions are not explicitly represented. Once reflection is available, it becomes possible for the child to start constructing explicit representations of what utterances actually mean and how they fit into the social world—i.e., what the presuppositions are of and how they modify situation conventions. This process takes time and effort, but by age 4, when reflection seems to emerge, the child has already constructed considerable knowledge of the linguistic realm of interaction, whose implicit presuppositions can be examined and represented. That is, content is there, but it is not as-of-yet represented explicitly.

More mature linguistic interaction, such as having a conversation about things that are not there, requires its participants to exercise reflection and to understand the meaning of utterances symbolically. That is, a toddler can have a conversation of that kind—e.g., about clouds and pets—but will be incapable of representing and considering in the conversation the abstract properties of those objects, such as the "hidden" causes of their behavior. In other words, language (i.e., situation conventions involving linguistic interaction) constitutes a realm of interaction that can be fully successfully navigated only with proper reflective understanding. As such, it imposes a selective pressure on the child's budding reflection—language-based interaction tests out the child's attempted constructions of reflective understanding and selects only those that afford successful anticipation of the interactive flow. Naturally, the child is aided in this developmental task by caregivers who engage in all kinds of functional scaffolding to lower the selective pressures inherent in language (Bickhard, 1992): Repeating things, narrating while demonstrating and so on. Language, then, is a realm of interaction that serves both as a motivator for reflective construction and as a testing ground for it. Without an opportunity to interact linguistically, reflective understanding is critically impaired, as the tragic cases of language deprived children attest (Fromkin et al., 1974).

Further, as success in linguistic interaction drives the child's reflective construction (once enabled by CNS developments), by the same token it imparts some organization onto the child's resulting reflective knowledge. Not only due to its formal properties such as syntax or morphology, but also in terms of associations, symbolic tropes, or generally speaking—ways of thinking—that abide in a given language or culture more broadly. Indeed, it is hard to imagine how an organism would show culturally-constituted reflection without a language scaffolding the process, and thus it can be difficult to disentangle properties of our reflective thinking that stem from its linguistic formatting and those that characterize reflection as such. Perhaps due to this entanglement, many scholars in history have declared thought to have a language-like structure (e.g., Fodor, 1975), which from the interactivist perspective amounts to misattributing properties of language to the nature of reflective thought as such.

It needs to be stressed that cultures and languages differ, and that they do so to some extent in terms of what kind of reflective abstraction is needed to enter them; this can be both in terms of types of content—like mental state concepts vs. behavioral concepts—or ways of thinking about some content—like theory vs. narrative. These differences in interactive realms likely lead to children from those cultures to exercise their reflection in accordance with them and thus do better on tests that presuppose competence in those terms. For instance, the explicit change of location False Belief Task (FBT) is passed at different ages depending on culture—in the West it is around age 4, but in Japan at 6+ (Naito, 2014). Whereas, multiple factors can be evoked to explain this difference, the specificity of the folk conceptualizations about the social world that dominates in the two cultures might be a significant one. As Naito argues, rather than a theory of mind, Japanese folk theory is that of relations between people. To be sure, both of these conceptualizations are true in the sense that they abstract real aspects of the social world—individuality and epistemic separateness in the former case, and the interconnectedness in the latter—but the difference in emphasis seems to lead to differing developmental trajectories in what is reflectively represented, which seems to be reflected in children's performance on socio-cognitive tests. The FBT arguably requires the child to have a clear reflective understanding of how perceptual contact of an individual mind relates to their knowledge of the world—the kind of reflective understanding that American children steeped in Western folk psychology would develop early and Japanese children would find rather foreign. However, things are different with other socio-cognitive tests, such as ones that involve aspects more aligned with the Japanese theory of relations. For instance, in one such task the object about which the protagonist of the FBT forms a false belief is changed from a

9 Empirically speaking, there does not seem to be evidence for development beyond level 4. However, the issue has not been directly investigated.

10 The term "symbolic" is usually understood in an encodingist way; here, instead, we mean it in the interactivist sense, as explicit representation of implicit presuppositions about what words refer to.

physical object (e.g., a toy) to a person who has promised to stay in one place rather than the other, but moves unbeknownst to the protagonist (Symons et al., 1997; Naito, 2014, p. 390). Japanese children seem to do better than their Western counterparts on that test, and when they are asked to motivate their answers, they tend to cite social obligations such as "he promised he'd be there" rather than individual epistemic states of the protagonist such as their mistaken belief.

Finally, once understood symbolically, language greatly facilitates reflective abstraction; that is, symbolic and systematic language provides a format that externalizes thought, which facilitates the climb up the knowing levels. The fundamental principle of interactivist knowledge formation is that only that which can be interacted with can be represented. For levels 1 and 2, the epistemic access is direct—level 1 interacts with the structure of reality, both physical and social, via the senses; and level 2 interacts with the organization of level 1 knowledge, via the physiological links in the CNS. This leaves the question of how reflection can climb beyond these two levels of representing—how to represent the implicit presuppositions of level 2 knowledge and higher?

Action involving level 2 reflection will leave a mark on the organization of level 1, both indirectly by influencing how the agent acts in the world and directly via internal thought. Consequently, the reorganized level 1 knowledge will come to involve some of the presuppositions of the reflective processes, which will make it possible for those presuppositions to be represented, leading to the emergence of level 3 knowledge—an explicit characterization of level 2 presuppositions.

While in principle, this loop of externalization and reflective abstraction could proceed indefinitely, having a symbolic system that provides an external systematic format for mental content greatly aids the process. Knowing processes that are put in language can be examined in terms of their presuppositions regardless of the level of reflection. As discussed by Campbell and Bickhard (1986), Aristotle's development of syllogistic logic forms an illustrative example here: He started to use abbreviations for names in syllogistic sentences, which later became variables in the general form—reflective abstraction of the logical properties of level 2 reasoning into an explicit representation of those properties. Once that happened, it became possible to examine the presuppositions of that abstracted framework and construct a representation of them as Aristotle's syllogistic calculus—level 4.

5.1 Language and developmental methodology

Thinking about how language operates for pre-reflective thought has implications for methodological design and interpretation of empirical results. In general, language does not operate for 3-year-olds as it does for 4- and 5-year-olds. This means that the same instructions or manipulations have different consequences for the two groups. For example, consider social learning research focused on testimony (Harris et al., 2012). The canonical version of the *trust* paradigm involves someone (mis)labeling familiar objects to induce (un)reliability in one of two informants. From the interactivist perspective, the nature of this manipulation is different for pre-reflective 3-year-olds than it is for 4- and 5-year-olds. For 3-year-olds, the mislabeling cannot be a *reliability* manipulation *per se*. Reliability is a reflective attribute that can only potentially be represented by around age 4. The manipulation clearly has consequences for 3-year-olds in terms of their informant preferences, but we would suggest that the proper interpretation of those preferences is in terms of 3-year-olds avoiding the "unreliable" informant rather than selecting the "reliable" informant. This would mean that trust research is more appropriately characterized as being about "mistrust" for children under age four. Further, a scientific explanation of the reasons for their (mis)trust can be modeled in ways that go beyond dispositional explanations about credulity and skepticism.

For example, consider testimony paradigms with a single informant who makes a claim that differs from the child's own experience (Ma and Ganea, 2010). In this situation, an object is placed in an occluded location. An informant then claims it is actually at a second location. Three-year-olds, but not older children, chose to rely on the informant's information over their own experience. The explanation for this is that 3-year-olds are overly credulous. However, other evidence suggests that 3-year-olds are overly skeptical (Woolley and Ghossainy, 2013). This raises two issues: (1) which characterization is accurate; (2) being credulous or skeptical does not explain behavior so much as it describes a tendency to behave a certain way. From the interactivist perspective, 3-year-olds "credulity/skepticism" are both a consequence of language as transforming social realities. In the case of credulity, the informant's claim transforms the 3-year-olds interactive characterization into one in which the object is indeed at the second location. This happens because they cannot yet evaluate the utterance separate from its transformative consequences. In the case of skepticism, the testimony applies for claims about contents for which the child does not have interactive experience (e.g., fantastical/historical characters). Accordingly, the utterance in such situations has too little interactive characterization to transform. This is like trying to manipulate physical objects that do not exist. Accordingly, reflection will be required to represent fantastical objects in the first place such that an utterance can then serve its transformative function.

6 Conclusion

The current proposal sought to critically evaluate extant models of the emergence of representation during development (both for *foundational* emergence as well as for *subsequent* emergence). It was concluded that the limitations of these models ultimately derive from their own development within an information-processing framework. Interactivism was introduced as an action-based alternative to information-processing and its specific models of representation (foundational emergence) and reflection (subsequent emergence) were presented. Implications for the model of reflection were discussed in terms of some empirics, thinking about consciousness, enculturation as a construction process on the part of the child, and the role of language in that process with some examples involving the sociality of theory of mind. A final discussion opened the door to considerations about how language may affect developmental

methodology and interpretation for preschooler with reflective vs. pre-reflective capabilities.

Author contributions

JA: Conceptualization, Writing – original draft, Writing – review & editing. RM: Conceptualization, Writing – original draft, Writing – review & editing. MB: Conceptualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We would like to thank Stephanie Carlson for helping us work through the open access process and two reviewers for their comments. We would also like to thank Ezgi Ozgan, Deniz Hasan, and Yağmur Esendemir for their comments on a prior draft.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Allen, J. W., and Bickhard, M. H. (2011). You can't get there from here: foundationalism and development. *Behav. Brain Sci.* 34:124. doi: 10.1017/S0140525X10002311

Allen, J. W., and Bickhard, M. H. (2013). Stepping off the pendulum: why only an action-based approach can transcend the nativist-empiricist debate. *Cogn. Dev.* 28, 96–133. doi: 10.1016/j.cogdev.2013.01.002

Allen, J. W., and Bickhard, M. H. (2018). Stage fright: internal reflection as a domain general enabling constraint on the emergence of explicit thought. *Cogn. Dev*. 45, 77–91. doi: 10.1016/j.cogdev.2017.12.005

Allen, J. W., and Bickhard, M. H. (2022). Emergent constructivism: theoretical and methodological considerations. *Hum. Dev.* 66, 276–294. doi: 10.1159/000526220

Allen, J. W., Çelik, B., and Bickhard, M. H. (2021). Age 4 transitions: reflection as a domain-general development for explicit reasoning. *Cogn. Dev.* 59:101091. doi: 10.1016/j.cogdev.2021.101091

Baillargeon, R. (2008). Innate ideas revisited: for a principle of persistence in infants' physical reasoning. *Psychol. Sci.* 3, 2–13. doi: 10.1111/j.1745-6916.2008.00056.x

Bickhard, M. H. (1978). The nature of developmental stages. *Hum. Dev.* 21, 217–233. doi: 10.1159/000271586

Bickhard, M. H. (1980). *Cognition, Convention, and Communication*. Westport, CT: Praeger.

Bickhard, M. H. (1992). Commentary on the age 4 transition. *Hum. Dev.* 35 182–192. doi: 10.1159/000277151

Bickhard, M. H. (2005). Consciousness and reflective consciousness. *Philos. Psychol.* 18, 205–218. doi: 10.1080/09515080500169306

Bickhard, M. H. (2009a). The interactivist model. *Synthese* 166, 547–591. doi: 10.1007/s11229-008-9375-x

Bickhard, M. H. (2009b). Interactivism: a manifesto. *New Ideas Psychol.* 27, 85–95. doi: 10.1016/j.newideapsych.2008.05.001

Bickhard, M. H. (2015). Toward a model of functional brain processes II: central nervous system functional macro-architecture. *Axiomathes* 25, 377–407. doi: 10.1007/s10516-015-9276-9

Bickhard, M. H. (2024). *The Whole Person: Toward a Naturalism of Persons*. Amsterdam: Elsevier.

Bickhard, M. H., and Terveen, L. (1995). *Foundational Issues in Artificial Intelligence and Cognitive Science: Impasse and Solution, Vol. 109*. New York, NY: Elsevier.

Campbell, R. L., and Bickhard, M. H. (1986). *Knowing Levels and Developmental Stages. Contributions to Human Development, Vol. 16*. Berlin: S. Karger.

Carey, S. (2009). *The Origins of Concepts.* New York, NY: Oxford University Press.

Chomsky, N. (1959). A review of Skinner's verbal behavior. *Language* 35, 26–58. doi: 10.2307/411334

Christopher, J. C., and Bickhard, M. H. (2007). Culture, self and identity: interactivist contributions to a metatheory for cultural psychology. *Cult. Psychol.* 13, 259–295. doi: 10.1177/1354067X07079881

Dewey, J. (1915). "The existence of the world as a logical problem," in *The Collected Works of John Dewey, 1882–1953. 37 Vols. The Middle Works. Vol. 8, 83–97* (Carbondale, IL: Southern Illinois University Press), 1967–1987.

Dewey, J. (1941). "Propositions, warranted assertibility, and truth," in *The Collected Works of John Dewey, 1882–1953. 37 Vols. The Later Works of John Dewey, 1925–1953 (168–188)* (Carbondale, IL: Southern Illinois University Press), 1967–1987.

Engel, A. K., Friston, K. J., and Kragic, D. (2016). *The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science, Vol. 18*. Cambridge, MA: MIT Press.

Fodor, J. (1998). *In Critical Condition: Polemical Essays on Cognitive Science and the Philosophy of Mind*. Cambridge, MA: The MIT Press.

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

Fromkin, V., Krashen, S., Curtiss, S., Rigler, D., and Rigler, M. (1974). The development of language in genie: a case of language acquisition beyond the "critical period". *Brain Lang.* 1, 81–107. doi: 10.1016/0093-934X(74)90027-3

Gershkoff-Stowe, L., and Thelen, E. (2004). U-shaped changes in behavior: a dynamic systems perspective. *J. Cogn. Dev*. 5, 11–36. doi: 10.1207/s15327647jcd0501_2

Gopnik, A. (2003). "The theory theory as an alternative to the innateness hypothesis," in *Chomsky and His Critics*, eds. L. Antony and N. Hornstein (New York, NY: Basil Blackwell), 238–254.

Harris, P. L., Corriveau, K. H., Pasquini, E. S., Koenig, M., Fusaro, M., and Clément, F. (2012). Credulity and the development of selective trust in early childhood. *Foundat. Metacogn.* 193–210. doi: 10.1093/acprof:oso/9780199646739.003.0013

Iao, L. S., Leekam, S., Perner, J., and McConachie, H. (2011). Further evidence for nonspecificity of theory of mind in preschoolers: training and transferability in the understanding of false beliefs and false signs. *J. Cogn. Dev.* 12, 56–79. doi: 10.1080/15248372.2011.539523

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

Karmiloff-Smith, A. (2018). "An alternative to domain-general or domain-specific frameworks for theorizing about human evolution and ontogenesis," in *Thinking Developmentally From Constructivism to Neuroconstructivism* (London: Routledge), 289–304.

Ma, L., and Ganea, P. A. (2010). Dealing with conflicting information: young children's reliance on what they see vs. what they are told. *Dev. Sci.* 13, 151–160. doi: 10.1111/j.1467-7687.2009.00878.x

Mandler, J. (2004). *The Foundations of Mind: Origins of Conceptual Thought*. New York, NY: Oxford University Press.

Mandler, J. M. (2000). Perceptual and conceptual processes in infancy. *J. Cogn. Dev.* 1, 3–36. doi: 10.1207/S15327647JCD0101N_2

Mirski, R., and Bickhard, M. H. (2021). Conventional minds: an interactivist perspective on social cognition and its enculturation. *New Ideas Psychol.* 62:100856. doi: 10.1016/j.newideapsych.2021.100856

Müller, U., and Kerns, K. (2015). The development of executive function. Handbook of child psychology and developmental science. *Cogn. Process.* 2, 571–623. doi: 10.1002/9781118963418.childpsy214

Müller, U., and Overton, W. F. (1998). How to grow a baby: a reevaluation of image-schema and Piagetian action approaches to representation. *Hum. Dev.* 41, 71–111. doi: 10.1159/000022570

Naito, M. (2014). "From theory of mind to theory of relation: sociocultural perspectives on Japanese Children's Social Understanding," in *Contemporary Perspectives in Early Childhood Education. Contemporary Perspectives on Research in Theory of Mind in Early Childhood Education*, ed. O. N. Saracho (Cambridge, MA: Information Age Publishing), 381–408.

Penn, D. C., Holyoak, K. J., and Povinelli, D. J. (2008). Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *Behav. Brain Sci.* 31, 109–130. doi: 10.1017/S0140525X08003543

Perner, J. (1991). *Understanding the Representational Mind*. Cambridge, MA: MIT Press.

Perner, J., Stummer, S., Sprung, M., and Doherty, M. (2002). Theory of mind finds its Piagetian perspective: why alternative naming comes with understanding belief. *Cogn. Dev.* 17, 1451–1472. doi: 10.1016/S0885-2014(02)00127-2

Piaget, J. (1952). Play, dreams and imitation in childhood. *Int. Libr. Psychol.* 1999:87.

Piaget, J. (1954). *The Construction of Reality in the Child.* New York, NY: Norton.

Piaget, J. (1971). Genetic epistemology. *Norton* 1971:piag91272. doi: 10.7312/piag91272

Piaget, J., and Inhelder, B. (2000). *The Psychology of the Child*. New York, NY: Basic Books.

Smith, L. (1993). *Necessary Knowledge: Piagetian Perspectives on Constructivism (Essays in Developmental Psychology)* (Hove: Lawrence Elrbaum Associates Ltd, Publishers), 155.

Symons, D., McLaughlin, E., Moore, C., and Morine, S. (1997). Integrating relationship constructs and emotional experience into false belief tasks in preschool children. *J. Exp. Child Psychol.* 67, 423–447. doi: 10.1006/jecp.1997.2416

Tiles, J. E. (1990). *Dewey*. London: Routledge.

Tomasello, M. (2024). An agency-based model of executive and metacognitive regulation. *Front. Dev. Psychol.* 2:1367381. doi: 10.3389/fdpys.2024.1367381

Turner, S. P. (1994). *The Social Theory of Practices: Tradition, Tacit Knowledge and Presuppositions/Stephen Turner*. Cambridge: Polity Press.

Turner, S. P. (2018). *Cognitive Science and the Social: A Primer/Stephen P. Turner (1st)*. London: Routledge.

Vygotsky, L. S. (1978). *Mind in Society: the Development of Higher Psychological Processes/L. S. Vygotsky; Edited By Michael Cole … [et al.]; [Translated From the Russian]*. Cambridge, MA: Harvard University Press.

Woolley, J. D. E., and Ghossainy, M. (2013). Revisiting the fantasy-reality distinction: children as naïve skeptics. *Child Dev.* 84, 1496–1510. doi: 10.1111/cdev.12081

Zelazo, P. D. (1999). "Language, levels of consciousness, and the development of intentional action," in *Developing Theories of Intention* (London: Psychology Press), 95–118.

Zelazo, P. D. (2004). The development of conscious control in childhood. *Trends Cogn. Sci.* 11:1. doi: 10.1016/j.tics.2003.11.001

Zelazo, P. D. (2015). Executive function: reflection, iterative reprocessing, complexity, and the developing brain. *Dev. Rev.* 38, 55–68. doi: 10.1016/j.dr.2015.07.001

The term *coding* along with *processing* and *computation* forms a trio of concepts whose ubiquitous use in neuroscience is underwritten by a usage so loose that often they can replace one another without appreciably affecting the sense of what is being conveyed, as in "We study how the brain codes (or processes, or computes) emotions." As placeholders for "something unknown or yet to be determined," these terms might seem innocuous, were it not for the fact that they carry with them semantic baggage from their nonneural provenance. This technical or formal sense figures in *some,* but far from all, of their uses in neuroscience discourse, a situation that promotes conceptual and communicative confusion.

Regarding the technical and formal sense of "code" and "coding" specifically, the core of the construct as used in the disciplines from which it was imported is that of rule-governed relations of correspondence between two domains with *arbitrary* correspondence assignments in the sense that alternative assignments would work. This is the only common denominator of the two principal codes to be found in nature: (1) the nucleotide-triplet code by which transfer RNA uses the base sequence of messenger RNA to string amino acids into proteins (other genetic operations being template-based), and (2) the two-level combinatorics of phonological and lexical elements by which the sound strings of human languages code the arbitrary (in this case conventional) mapping between the form and meaning of utterances. The arbitrary/conventional element is conspicuous also in the various artificial codes created in reliance on human language: ciphers in cryptography, Morse code in telegraphy, and the many coding schemes employed in the design of digital computers and their programs.

Bona fide codes represent a vanishingly small portion of the myriad lawful relationships that make up the natural world. Likewise for the nervous system: All neural operations of which we have actual knowledge are lawful ones lacking the arbitrary aspect of the correspondence rules of a coding scheme. There is no dearth of appropriate terms for such noncoding lawful relations: "function," "transfer function," "transduction," "mapping," "transform," "representation," and more. Yet the technically incorrect "coding" often substitutes for more informative terms, especially when incompletely known aspects of function, and issues of significance or meaning in particular, are being addressed.

The tacit analogy appears to be the message passing made possible by human language: Neurons "communicate," and somehow the nervous system generates meaning, so perhaps neurons send language-like messages to one another, coding significance or meaning in the temporal sequence of their spike trains. This would supply neurons with an extra-local "code for information" or common language.

But consider the roughly 8000 synapses impinging on a single cortical pyramidal cell: They originate in many hundreds, if not thousands, of other pyramidal cells and subcortical neurons. The blending at the axon initial segment of the graded potentials induced by the synaptic activity of all these sources, excitatory and inhibitory, jointly determines whether that cell will reach threshold to release its all-or-none action potential to its audience of hundreds or thousands of other cells. "Messages" do not survive such treatment anymore than it is possible to monitor the conversations of a cocktail party by a single decibel meter rigged to issue a spike at a fixed sound pressure level. The fact that close to threshold a situation can arise in which the spike output of a neuron replicates the firing of one or a few afferents does not change the basic irrecoverability of the pattern of afferent input from a neuron's output.

Regarding significance, more than half a century of arduous mapping of the response properties of single neurons throughout the nervous system tells us that the principal determinant of what the activity of a neuron signifies is *where it is located*. This "where" ranges from the gross subdivisions of the brain down to a cell's precise connective position in the synaptic network it inhabits (Passingham et al. 2002). Moving a microelectrode to a neighboring cell typically discloses slightly different, but related, response properties, in the aggregate generating the familiar functional maps that abound at every level of the neuraxis.

The reason for this parallel and analog mode of representing significance in the brain is not far to seek. The cortical signal propagation speed of 1 to 6 meters per second (based on Pascual-Leone et al. 2000; Schmolesky et al. 1998) is some 180 million to 30 million times slower than that of electronic circuitry. Operations in this sluggish medium must deliver their global results some three to four times per second (the frequency of gaze movements, the leading edge of most behavior; Merker 2013b). The brain is therefore always strapped for time, and as computer programmers know, when time is short, you don't compute, you use lookup tables.

To perform sophisticated functions with sluggish components in real time, the brain arrays them in complex concatenations of innate as well as acquired (learned) maps ("lookup tables" for "priors") interfaced with one another via a variety of connectivity-based functional transformations. Together they form an analog "computer," not a digital one, for which the analog inner workings of so-called neural networks supply toy models. In analog computers there is no program running on the hardware: The hardware itself *is* the program, and that hardware, moreover, is modifiable by its own activity, generating the learned content of maps.

With no program to run, and no messages to send, there is nothing to code, because significance is not represented in propositional or symbolic form in the brain, but positionally: *where* activity flares is what it signifies or means. But what about the signal discretization of the action potential? A neuron's work takes place by analog blending of graded potentials on its soma-dendrite membrane, and spikes serve only to transmit with fidelity a running record of the upshot of that work to other locations, in keeping with Shannon's insight regarding the utility of discretization for faithful signal *transmission*. No neural work is being performed by discrete spikes except that of bridging distance.

In sum, the answer to the question posed in Romain Brette's title is a resounding "No!"

Encodingism is not just a bad metaphor

Robert Mirski[a,b] and Mark H. Bickhard[b]

[a]The John Paul II Catholic University of Lublin, 20-950 Lublin, Poland and [b]Lehigh University, Bethlehem, PA 18015.
robertmirski@kul.lublin.pl
mhb0@lehigh.edu
https://www.researchgate.net/profile/Robert_Mirski
https://www.lehigh.edu/~mhb0/mhb0.html

doi:10.1017/S0140525X19001286, e237

Abstract

Brette's criticism of the coding metaphor focuses on its presence in neurosciences. We argue that this problematic view, which we call "encodingism," is pernicious in *any model of cognition that adopts it*. We discuss some of the more specific problems it begets and then elaborate on Brette's action-based alternative to the coding framework.

Brette argues that encodingism assumptions are pernicious in neuroscience. We would like to expand this critique a bit: Encodingism is a problem in models of cognition in general, not only in neuroscience. We argue that, though encodings certainly exist, they are derivative by nature, and cannot serve to explain the basis of natural cognition. As Brette points out, neurons *could* be said to "encode" the information about some property *for neuroscientists*, as it is they who are interpreting the coding relationship. That is, encodings always require an interpreter who already knows about or represents the two ends of the encoding relationship, as well as the relationship itself. But this representation is exactly the knowledge we are trying to account for when researching minds, and so encodingism becomes circular and leads to an infinite regress of interpretive homunculi. Something else has to lie at the bottom of the natural ability to represent.

The above point underlies Brette's article, but it is also important to note that there is a whole family of problems that plague encodingism. Some of these problems have withstood resolution for millennia. For example, the impossibility of system detectable error that Brette mentions can be traced back to classical scepticism – how am I supposed to know whether what I represent is true, if, in order to find that out, I would have to step outside of myself to gain some independent epistemic access and check? The other end of an encoding is, supposedly, some entity, or property, or state of affairs, but if encodings are all the system has available to represent its reality with, then the only way to attempt to check the encoded end of an encoding is use another encoding. Circularity again.

Foundationalism is another problem forced by encodingism assumptions. Within an encodingist framework, it is impossible for the organism to create first encodings or representations for the very same reason stated above – the organism would have to already know what this particular information is "about" to use it to create a representation. Circularity for a third time.

One would think that this impossibility of representational emergence should automatically discredit encodingism among developmentalists who study the origins of mentality. However, this has not always been the case. Rather, the problem of emergence has been pushed onto biology, and various "core knowledge" accounts have been proposed: infants are supposed to be born with innate theories of physics, biology, or mind (for more criticism, see Allen & Bickhard 2013; Mirski & Gut 2018). But if encodingism blocks emergence in ontogeny, there is no reason why it would not do so in phylogeny too. These are just three of many more problems; for more, see Bickhard and Terveen (1996; Bickhard 2009).

What alternatives are there then? Brette's proposal that we should ask what neurons do rather than what they encode is a significant step in the right direction. However, there are further aspects of cognition, which Brette does not discuss, that we would like to briefly address. Organisms certainly represent reality, and can be wrong about it, and when they are wrong, they often discover that and learn from their mistakes. Naturalism requires that whatever constitutes this representing, and representational error detection, has to emerge at some point from non-representational phenomena. As has been argued, none of these can be accounted for in encodingism in principle, but an action-based perspective has to provide an alternative on pain of being explanatorily vacuous.

Brette briefly mentions what we take to be central to an action-based model when he says "what is useful for the organism is not literally to predict what will happen next, but rather what *might* happen next, conditionally on the actions I can do, so that I can select the appropriate action" (sect. 3.4, para. 6). This statement contains a hint of what mental content can be in an action-based model – the anticipation of possible interactions. This is the proposal of interactivism (Bickhard 2009). (Strictly, it is the anticipation of possible internal process flows that are co-determined by the environment and the organism's actions; it is not anticipation of interaction with the environment as such – there is no surview of the organism and its environment. See, for example, Bickhard 2009; 2015a; 2015b). Such anticipations will have truth value – they implicitly predicate something about the environment (i.e., it is the kind of environment that supports this kind of interaction). And they will be in principle falsifiable and detectable by the organism – all it takes to see if I am right is to actually (try to) engage in the interaction.

As for learning and initial emergence of such action-based normativity, it can happen if we adopt a variation and selection model of learning. If successful anticipations are retained and unsuccessful anticipations are selected against, then the limiting case of representational emergence will be to randomly engage in various interactions and retain the ones that turn out to be successful. No prescience is necessary like in encodingism. Similarly in learning, if my anticipations are falsified, I vary the way I do things until I stumble on a successful alternative. A more detailed discussion of these points can be found elsewhere (Bickhard 2001; 2003; 2009; 2015c; Bickhard & Campbell 1996).

Conceiving of brain functioning in terms of such an anticipatory organization is a viable alternative to coding (Bickhard, 2015a; 2015b). On this view, the brain establishes modes of functioning that implicitly anticipate the upcoming interaction. The modes of functioning are set up by the modulations of such elements as volume transmitters, astrocytes or silent neurons. Such modulations *are* anticipatory in that they set particular modes up, which could turn out to be inappropriate modes for what process flow actually happens. Adopting this alternative, anticipatory view of the brain could complement and extend Brette's proposal. (The above has similarities to some other contemporary frameworks, especially to predictive processing [Clark 2016] and enactivism [Di Paolo & De Jaegher 2012], and indeed there are considerable overlaps, but also fundamental differences [Bickhard 2015b; 2016a; 2016b]).

Acknowledgments. RM was supported by a grant from the National Science Centre (UMO-2016/23/N/HS1/02887).

Michael 2015), is that infants and young children progressively refine their agent-models through interaction with others, and also increasingly conform to those models themselves. More concretely, infants and young children apply agent-models to others during the course of development, and use these agent-models to guide imitation and other forms of cultural learning. A result of this is that it becomes increasingly feasible for others in their culture to model them as well. In other words, agents' models of themselves (their self-models) and their actual selves are fitted together as a natural consequence of modeling and interacting with each other. From the perspective of the prediction error minimization framework, this appears as a form of active inference: infants and young children shape their selves progressively to match the agent models that they have been using to interpret others.

We hope that careful consideration of these three points will help in clarifying the relationship among culture, theory of mind, and predictive coding, and that it will stimulate progress in explaining the dynamics by which minds shape, and are shaped by, other minds through the complex process of enculturation.

Enculturated minds, not error reduction minds

Robert Mirski[a], Mark H. Bickhard[b], David Eck[c] and Arkadiusz Gut[d]

[a]The John Paul II Catholic University of Lublin, 20-950 Lublin, Poland; [b]Lehigh University, Bethlehem, PA 18015; [c]Cañada College, Redwood City, CA 94061 and [d]The Nicolaus Copernicus University in Toruń, 87-100 Toruń, Poland.
robertmirski@kul.lublin.pl https://www.researchgate.net/profile/Robert_Mirski
mhb0@lehigh.edu https://www.lehigh.edu/~mhb0/mhb0.html
eckd@smccd.edu https://canadacollege.academia.edu/DavidEck
arekgut2001@gmail.com https://umk.academia.edu/ArkadiuszGut

doi:10.1017/S0140525X19002826, e109

Abstract

There are serious theoretical problems with the free-energy principle model, which are shown in the current article. We discuss the proposed model's inability to account for culturally emergent normativities, and point out the foundational issues that we claim this inability stems from.

We believe the free-energy principle (FEP) lacks theoretical resources to account for the complex phenomenon of culture. The current article's attempt at doing so results in a trivialization of the problem, and a reductionist view on what culture and its participants are. Below we focus on the problems the proposal faces with accounting for the diverse normativities that characterize enculturated persons. After that, we argue that this is a symptom of more fundamental theoretical problems with the FEP.

The FEP claims that the overarching goal of every individual is to reduce free energy or uncertainty. Accordingly, all normativities that the system instantiates are claimed to come from the preselected set of "expectations"; for instance, living organisms are argued to move away from dangerous temperatures because these temperatures generate inputs incompatible with "expectations" about them (this is the example given in the current article). These adaptive "expectations" are argued to reside in the highest level "expectations," sometimes called hyperpriors (Clark 2013a), which have been formed during phylogeny; only those individuals with adaptive hyperprior "expectations" managed to survive and procreate (Friston et al. 2012; Kiebel et al. 2008).

Although a rather ingenious idea, the above claim runs into clear problems in the context of enculturation. People certainly have phylogenetically old normativities such as the ones satisfying our basic survival needs, but they also house a whole plethora of normativities emergent over the course of development, ones that cannot be argued to have formed in phylogeny. It hardly needs demonstration that genetically identical and raised in the same socio-cultural milieu twins can develop radically opposing sets of values and goals. What is more, these goals and values can sometimes override the phylogenetically old, adaptive normativities: history knows many cases of people deciding to die or suffer for some highly abstract cause. This fact seems entirely incompatible with the FEP model, and it is especially problematic in the context of the current proposal because these powerful, novel normativities usually emerge as part of the process of enculturation. In fact, encultured persons *are constituted* by such emergent normative phenomena: We certainly can identify more with our values and goals than with our biologically given motivation to stay alive, which itself is far from defined innately as it emerges ontogenetically in a social context too (e.g., we learn the "proper" ways of eating or sleeping from our cultures) (see Eck & Levine 2017).

In the context of the multi-layered human cognitive system, the highest-level, adaptive normativities given in hyperpriors are argued to yield information-seeking or global-uncertainty-reduction dynamics. This is held up in the current article as solving the "dark-room problem": increases in local uncertainty are expected to decrease global uncertainty over time, that is, to keep the organism within the innately expected states specified in the hyperpriors. This claim seems to give us another kind of normativity that is derived from the overarching motivation of the FEP: namely, the epistemic-gain motivation. Unfortunately, this does little to help the situation as motivations emergent in encultured persons cannot be reduced to information seeking either. How does my re-watching for a hundredth time an old cult movie at my house benefit me epistemically? In fact, culturally emergent normativities are sometimes flatly hostile to epistemic gain – ignorance passes for cool in some communities.

These issues of the FEP being incompatible with the reality of culturally emergent normativities bear heavy on the proposed model. Although the paper talks about relevant phenomena – such as norms, affect, or prestige – as if they have been explained (there are many such glaring cases of petitio principii in the article), the proposed model hinges solely on the epistemic-gain motivation. Culture boils down to informational redundancies created for a more efficient epistemic gain (cultural niche construction) and individuals learning about these redundancies (learning cultural affordances). There are no persons with their ontologically novel normativities, such as values, ideals, and other diverse motivations – just individual organisms helping each other minimize their free energy (for an analysis of the problem of such "manipulationist" views on culture in other models, see Eck 2015).

It becomes clear that the predictive processing framework and the FEP are not fit for modeling culture. Following Litwin and Miłkowski (submitted), we believe that the framework needs serious theoretical development before it can be fruitfully applied to specific problems such as the one at hand. Indeed, the inability to model normative emergence in enculturation is an important

special case of a more general problem: Free energy cannot handle *any* normative phenomena per se – neither the ones involved in enculturation nor the ones involved in development in general, not even the basic normativities inherent in life (Bickhard 2015; 2016; Martyushev 2018; Roesch et al. 2012). At best, pre-programed hyperpriors can extensionally capture predetermined behavior patterns. Any exceptions (e.g., seeking dark, instead of turning on the light; seeking pain [e.g., hot peppers] instead of avoiding it; etc.) must also be pre-programed: there is no modifying the hyperprior probabilities (they are innate, as are *all* of the *spaces* over which *all* of the probabilities are distributed) – there is no normative learning, no development, no socialization, and no enculturation, the last of which we discussed in this commentary. For a related discussion of problems with such foundationalism in cross-cultural research, see Mirski and Gut (2018). For an anticipatory framework that does address issues of normativity, see, for example, Bickhard (2009) and Campbell (2015) – including in the context of culture and language (e.g., Bickhard 1992; 2007; 2008).

Acknowledgments. RM was supported by a grant from the National Science Centre (UMO-2016/23/N/HS1/02887).

Importance of the "thinking through other minds" process explored through motor correlates of motivated social interactions

Harold Mouras

Laboratoire de Neurosciences Fonctionnelles et Pathologies (LNFP) EA 4559, Université de Picardie Jules Verne, Amiens, France.
harold.mouras@u-picardie.fr
https://www.u-picardie.fr/m-harold-mouras-202138.kjsp

doi:10.1017/S0140525X19002656, e110

Abstract

We wanted to gather recent results supporting the idea of the central role of sharing agency in socioaffective and motivational information processing. Here, we want to support the idea that this process is quite arbitrary, early in the temporal chain of processes and not only influence the psychological, but also the motor correlates of socioaffective information processes.

In their target article, Veissière et al. provide new theoretical arguments supporting the idea that "the human sense of obligation is intimately connected with the formation of a shared agent 'we', directing collaborative efforts and self-regulating them." Thus, they argue that "the human sense of obligation may thus be seen as a kind of self-conscious motivation."

Recently, several studies have brought experimental arguments supporting this idea by showing that the cognitive processes involved in this "formation of a shared agent" directly influence the psychological and motor correlates of socioaffective processes. Mainly, these studies have used a well-known theoretical and experimental model, that is, empathy for pain, which compares the processes (ratings, motor correlates) when viewing painful situations as compared to non-painful situations. Classically, the difference between the two conditions can be used as an index of empathy felt toward the character involved in the depicted situation. Therefore, several studies have been able to manipulate the nature of the social link between the observer (i.e., the participant) and the depicted character, in order to explore its influence on empathy.

Within social psychology, it is well-known that people have the propensity to divide the social world into *us* versus *them* influencing affective, cognitive, and behavioral processes. Interestingly, a powerful old paradigm (the minimal group paradigm; Tajfel et al. 1971) had demonstrated that the mere categorization of individuals into two social groups on the basis of arbitrary criteria (e.g., to over- or under-estimate the number of dots on a screen; Diehl 1990) was sufficient to produce similar consequences as compared to natural groups. We used for the first time this paradigm within the framework of empathy for pain (Montalan et al. 2012). Briefly, participants were shown pictures of people in painful or non-painful situations and were instructed to imagine themselves or imagine members of two minimal groups (in-group vs. out-group) in the same situation and participants had to rate the level of perceived pain according to the different perspectives. The results were quite clear: More than replicating previous results showing that the mere assignment of individuals to arbitrary groups elicits evaluative preferences for in-group relative to out-group members (Brewer 1979), we found that the mere act of categorizing people in two distinct social groups was also sufficient to elicit an in-group bias in empathy for pain. This was the first clear demonstration that the processes involved in the formation of a shared agent mentioned as central in Veissière et al.'s target article influenced the psychological processes of empathy.

What about the motor processes involved in socioaffective responses? We have been able to address this question by measuring the postural correlates of empathy for pain. The interrelation between the motor and affective components of behavior has been studied for a long time. For some theoretical models, emotion shapes behavior so that pleasant events should trigger approach whereas unpleasant events should trigger withdrawal. The ability to simulate another person's emotional response in a particular situation could be the basis for the development of empathic skills (Meltzoff & Decety 2003) and the instruction to adopt another person's perspective modulates pain rating according to the affective link between the observer and the individual experiencing the outcome (Singer et al. 2006). In a first study (Lelard et al. 2013), we used posturography to record differential postural responses when participants were instructed to imagine themselves in a painful or non-painful situation within the functional context of empathy for pain. This study demonstrated for the first time a stiffening response to pain visual stimulation, showing that postural responses were dependent of the perceived pain during the induced simulation process. These results laid the basis for further studies the basis for further studies concerning the role of perspective-taking in motivational dimension of motor control and social interaction. However, a main limitation of this study was that the effects of mental simulation were not tested, being unable to determine whether the reported effects were because of embodiment of the situation or to the valence of the visual scene.

A second study (Lelard et al. 2017) was designed to record the differential postural correlates of empathy for pain according to whether or not participants were instructed to imagine themselves in a painful or non-painful situation. Both painful visual scenes