Under Pressure: Processing Representational Decoupling in False-Belief Tasks

Anna Ciaunica

Abstract Several studies (Onishi and Baillargeon 2005; Surian et al. 2007) demonstrated that children younger than 3 years of age, who consistently fail the standard verbal false-belief task, can anticipate others' actions based on their attributed false beliefs. This gave rise to the so-called "Developmental Paradox". De Bruin and Kästner (2012) recently suggested that the Developmental Paradox is best addressed in terms of the relation between coupled (online) and decoupled (offline) processes and argued that if enactivism is to be a genuine alternative to classic cognitivism, it should be able to bridge the "cognitive gap", i.e. to provide us with a convincing account of how low-level sensorimotor practices transform into higher-order representational skills. This paper defends, against De Bruin and Kästner, an enactive response to the Developmental Paradox. I argue that 3-year olds' failure to verbally report their false-belief understanding does not arise from stronger decoupling demands. Rather, they fail because the elicited response false-belief trials involve representational decoupling *tout court* and what is more, under pressure.

1 Introduction: The Cognitive Gap

There is a new gap in town, namely the cognitive gap. Unlike the mind-body problem and its famous inheritor, the explanatory gap, the cognitive gap challenge is supposed to build a bridge the between low-level sensorimotor embodied practices and higher-order representational skills. Indeed, it has been argued that when it comes to "high-level" cognition, the representationalist paradigm is still without serious rival. Representationalism is roughly the view that "the manipulation and use of representations is the primary job of the mind" (Dretske 1995: xiv). Recently, representationalism has been challenged on the grounds that we have evolved from creatures whose neural resources were devoted primarily to perceptual and motoric processing, and whose cognitive activity consisted largely of immediate interaction with the environment (Bermúdez 2003; Chemero 2009; Gallagher 2005; Hutto 2008; Hutto and Myin 2013; Menary 2007). These theorists assume that cognitive processes are underpinned

A. Ciaunica (✉)
University of Fribourg, Av. de l'Europe 20, CH-1700, Fribourg, Switzerland
e-mail: ciaunica@yahoo.com

by the same processes that are used in physical interactions. Cognition is embodied in the sense that it depends on the experiences that result from possessing a body with given physical characteristics and a particular sensorimotor system. Indeed, a growing number of researchers in cognitive science are proposing that rather than viewing the body as a support system for a "mind that need to be fueled and transported" (Pecher and Zwaan 2005), one needs to view the mind as a support system that facilitates the functioning of the body. The embodied/enactive cognition (EC) approaches split in turn into two branches, (i) reconciliatory and (ii) radical, depending on their commitment to the idea that basic cognition involves or not the manipulation of representations. On the reconciliatory side, some theorists call on representations to explain behaviour, although they call on them in such a way that the need for internal manipulation of symbols is reduced. The representations they call on are either indexical-functional (Agre and Chapman 1987), "pushmi-pullyu" (Millikan 1995), action-oriented (Clark 1997), or emulator representations (Grush 1997; Churchland 2002). Opposing this view, radical embodied cognition (REC) challenges representationalism's conceptual foundations, namely the idea that mentality is essentially content involving and an exclusively heady affair. This relatively new research paradigm in cognitive science aims to study perception, cognition, and action as necessarily "embodied phenomena, using explanatory tools that do not posit mental representations. It is cognitive science without mental gymnastics" (Chemero 2009:29, emphasis added). Of course, this is not to deny the existence and importance of contentful and representationally based forms of cognition; it is just that these should be regarded as emerging late in phylogeny and ontogeny, being dependent on immersion in specific sorts of shared interactions (Hutto and Myin 2013).

When it comes to basic forms of cognition, the infant is, to use Wordsworth's famous metaphor, "father to the man". The issue at stake is crucial since as Sommerville and Woodward pointed out, assessing infants' understanding of others' behaviour provides not only a "snapshot of the developing mind of the child, but also a panorama of the very nature of cognition itself" (2010:84). Several studies (Onishi and Baillargeon 2005; Scott et al. 2010; Surian et al. 2007) demonstrated false belief understanding in toddlers via completely nonverbal measures such as the violation of expectation looking paradigm [1] (VOE henceforth). This gave rise to the so-called "Developmental Paradox" (DP henceforth): if false belief understanding already emerges during the second year of life, then why do 3-year-old children consistently fail the elicited-response task[2] (ERT henceforth) (Wimmer and Perner 1983; Wellman et al. 2001) even when paradigms are used that reduce response selection and inhibition demands (Sodian et al. 2006)?

It seems then that radical enactivists have to meet the following challenge:

P1   Empirical evidence strongly suggests that basic cognition is smart (since 15-month olds seem to understand false-beliefs).

---

[1] The VOE protocol tests whether children look longer when agents act in a manner that is inconsistent with their false beliefs and relies on the basic assumption that when an individual's expectations are violated, she is surprised and thus she looks longer at an unexpected event rather than at an expected event.

[2] In this famous experiment, dolls are used to act out a scenario in which Sally hides a marble in a basket and leaves the room. While she is gone, Anne enters and moves the marble to a box. Sally returns, and the children are asked, "Where will Sally look for her marble?"

P2   Smart cognition necessarily involves computations and representations (of false beliefs).
P3   Hence, basic cognition necessarily involves computations and representations (of false beliefs).

There have been roughly two main representationalist lines of response to the Developmental Paradox. Some authors argued that preverbal human infants do represent/understand false beliefs and explain the puzzle by arguing that ERTs require not only the ability to ascribe false beliefs, but also language understanding, pragmatic competence and more importantly, executive functioning control.[3]   Call this the rich interpretation (Baillargeon et al. 2010; Hala et al. 2003; Leslie et al. 2005).

Another solution is to account for P1 without crediting infants with false-belief understanding. The idea is to moderate P2 by claiming that preverbal infants' understanding is either behavioural rule-based (Perner and Ruffman 2005) or supported by an efficient but inflexible subsystem manipulating contentful informational states (Apperly and Butterfill 2009). Indeed, some theorists claim that apparent success on VOE trials in early infancy can be explained without supposing that infants have an understanding of beliefs at all. Call this the weak interpretation.

Opposing these representationalist views, radical enactivists aim to account for infants' success on VOE trials without endorsing P2 at all. They argue that one could simply let go of the idea that perceptions are representations and that basic cognition is essentially contentful in the first place. Rather, "basic cognition is literally constituted by, and to be understood in terms of, concrete patterns of environmental organismic activity, nothing more or less" (Hutto and Myin 2013:11).

Recently, De Bruin and Kästner (2012) proposed a middle-ground solution between representationalism and enactivism, namely Dynamic Embodied Cognition (DEC) and argued that when it comes to social cognition, "current articulations of enactivism are—despite their celebrated successes in explaining some cases of social interaction—not yet up to the task"(2012:541).

The aim of this paper is to illustrate that endorsing a reconciliatory-based solution does not provide us with a satisfactory solution to the Developmental Paradox. Rather, I shall claim, the puzzle is better explained within a REC-based framework, i.e. in terms of direct perception and situated interactions between agents and worldly offerings. First, I argue that some cognitivist residuals prevent DEC from truly meeting the challenge of the Developmental Paradox. Then I suggest an enactive response to the DP and argue that two implicit ingredients of the ERT tests, namely i) the verbal interaction and ii) the time–pressure factor, have been disregarded although they play a crucial role in false-belief failures in 3-year old children. This paper proceeds as follows. In section 2 I briefly sketch the reconciliatory Dynamic Embodied Cognition (DEC) account. In section 3 I address the issue whether the DEC is more likely to succeed in explaining the DP. Finally, section 4 illustrates how a REC-inspired strategy

---

[3] Executive functioning refers to a set of cognitive faculties that underlie goal-directed behavior and cognitive control across conceptual domains such as inhibitory control – i.e. the capacity to overcome one's putatively prepotent tendencies to simply say what is true and known – working memory, attention and error monitoring, etc.

meets the challenge of explaining the DP without crediting preverbal infants with sophisticated metarepresentational decoupling.

2 The Dynamic Reconciliatory Strategy: Taking Enactive Approaches on Board

In recent years, the propensity to analyse cognition on the basis of purely abstract models totally unrelated to the life of the body, of the social/physical environment, and of the brain regions governing the body's functioning in the world, has been vigorously challenged. As a result, a number of theorists have agreed to take certain enactivist-based insights on board, while still maintaining a cognitivist-based explanatory framework. For example, De Bruin and Kästner (2012) recently proposed that the Developmental Puzzle is best addressed in terms of the relation between coupled (online) and decoupled (offline) processes for basic and advanced forms of (social) cognition as opposed to merely representing/ not representing false beliefs. In what follows, I examine whether their middle-ground strategy is more likely to explain the puzzle. The upshot is to illustrate that although the DEC interpretation proposes some very interesting and helpful insights, the Developmental Puzzle remains unsolved.

Before delving deeper into De Bruin & Kästner's argument, some terminological clarifications are needed. They define cognition[4] as a kind of "information processing, without necessarily involving symbolic representations per se" (2012:541). Building upon this minimal definition, De Bruin and Kästner hold that the main characteristic of offline processing is the manipulation of information that is currently unavailable (i.e. not directly observable) in the environment (but typically has been present on previous occasions and therefore has to be internally represented). In contrast, online processing does not require recourse to such internal representations, but instead manipulation of information that is currently and readily available (i.e. directly observable) in the environment. Representations can be "any kind of stand-in for another item that will typically be understood as grounded in enaction but subsequently taken offline" (De Bruin and Kästner 2012:553, emphasis added). Note that taking processes offline can provide a serious advantage because it allows the agent to withdraw from the immediate surroundings so as not to automatically act upon particular affordances. The DEC is set up to accommodate both (1) the role of interaction with one's environment (enactivist-based thesis) and (2) the role of offline processes in social cognition (cognitivist-inspired thesis). Although the authors acknowledge a substantial cognitivist heritage insofar as they conceive of cognition as information processing, they nonetheless insist on the importance of bodily actions and environmental features in at least some cases of online cognitive processing. However, they go on arguing that if enactive/embodied accounts are to bridge the cognitive gap, they must not only acknowledge advanced offline social cognitive abilities but also explain these "decoupled" modes of social cognition without falling back into the cognitivist camp.

---

[4] One anonymous referee pointed out that De Bruin & Kästner could adopt another definition of cognition while keeping the fundamentals of DEC intact. However, as we shall see shortly, it is crucial for their solution to the puzzle to hold that preverbal infants are already capable of processing – via offline decoupling – internally represented information. REC challenges this claim and argues that situating the decoupling processing earlier in development does not help us to bridge the cognitive gap.

That is, "the enactivist has to tell a story about how offline (i.e. decoupled) social cognition is grounded in and emerges from online (i.e. coupled) interaction" (2012:547).

According to De Bruin and Kästner, rephrasing the issue in terms of online/ offline processing provides us with an explanation of the Developmental Puzzle. How exactly does this work? First, the authors take for granted the premise that infants are equipped with "implicit" abilities that start out as grounded in basic online processes, albeit partly decoupled. It is crucial for their project that these basic implicit abilities already involve decoupling. This is in line with the cognitivist distinction between (a) sub-doxastic mental states that do not possess truth-evaluable propositional content and (b) robust mental states (Spaulding 2010:123). In a second step, they hold that infants' "implicit" abilities develop gradually into more sophisticated explicit abilities that rely on offline processes to a much larger extent. Here it is worth quoting De Bruin and Kästner at length:

"Cognition is (developmentally) grounded in the agent's coupling to her surroundings. This coupling is dynamic insofar it allows agents to (a) rely on direct online cognitive processes, or (b) decouple from their environment and engage in offline cognitive processing. Online processing is 'cheap' and efficient as it allows the agent to avoid building up 'costly' internal representations which would require additional processing resources; however, it comes at the cost of limited and inflexible responsiveness. Offline processing provides agents with more flexibility (autonomy) regarding their direct environment, but is also cognitively more demanding. For a system to engage in decoupled offline processing, it has to (a) inhibit automatic behavioral responses to environmental features and (b) be able to elicit the behavior formerly coupled to the presence of a certain stimulus independently of environmental stimulation." (2012:553).

The coupling and decoupling relations between agent and environment advocated by DEC are dynamic in the sense that "they are a matter of degree and never an end in itself. (…) The dynamic interplay of decoupled and coupled processes may be used for optimization of cognitive processing." (De Bruin and Kästner 2012: 552 emphasis added). There is definitely much more to be said about DEC, but this gives us the basic flavor. With these elements in mind, let us examine whether this helps us to solve the DP.

3 Cheap Decoupling is not Enough: the Puzzle is Still There

Clearly, DEC borrows from the weak-strategy theorists such as Apperly and Butterfill (2009) the idea that early mechanisms are "cheap" and "efficient", while the late-emerging mechanisms are "costly" but "flexible". But they also borrow from rich theorists (Baillargeon et al. 2010) the idea that preverbal human infants are already capable of decoupling, i.e. taking their own reality-congruent perspective offline, albeit

in a very limited way. Their ambition is to provide us with the missing link between these two mechanisms without falling into cognitivist and enactivist extremes. The necessary glue is supplied by the Dynamical System Theory,[5] which is best suited to explain the architecture of the dynamic interplay between coupled and decoupled processes.

There are several problems with this interpretation of the Developmental Paradox. First, De Bruin and Kästner argue that VOE scenarios require infants to process differences between the visual information available to them alone and the visual information available to the other agent, and this can only be accomplished offline, since the other's visual information is not directly available to the infant and needs to be internally represented by her. Hence they jump to the conclusion that VOE trials involve a capacity for (i) decoupling from one's own online processing of visual information and (ii) processing offline a representation of the visual information accessible to another agent. This claim, however, is rather controversial. The other's mental state, although not directly observable (from a third-person perspective) to another, remains directly perceivable and available in a second-person perspective. The next section will make all of this clear.

Secondly, De Bruin and Kästner argue that ERTs require infants to not only represent but metarepresent. That is, infants must not only (a) process offline the other agent's propositional attitude towards the object (i.e., her false belief about its location), but also "(b) come up with a verbal prediction of the agent's behavior on the basis of her belief" (2012: 558). They claim that the decoupling required in (a) has been demonstrated to place increasing demands on executive functioning and they support their claim by quoting several studies stressing the robust correlations between: i) ERT performance and response inhibition (Perner and Lang 1999; Cole and Mitchell 2000; Carlson and Moses 2001); and (ii) ERT performance and working memory (Carlson et al. 2002; Hala et al. 2003; Perner et al. 2002). Again, this is disputable. Evidence from cross-cultural studies (Chen et al. 1998; Ho 1994; Sabbagh et al. 2006) and developmental cognitive neuroscience (Saxe and Powell 2006; Ridderinkhof et al. 2004) casts doubt on the putative relation between executive function and false belief performance. Indeed, in East Asian cultures, parents and teachers insist on the social-ization of self-control in their preschoolers (Chen et al. 1998; Tobin et al. 1989). Given this particular socialization, two hypotheses have been tested: (1) cross-cultural differences in the neurotransmitter systems will affect frontal lobe development and executive functioning performance (Chang et al. 1996); and (2) children from East Asia may have earlier developing executive functioning skills relative to their Western peers. Now, if Chinese preschoolers indeed have advanced executive skills, then according to the executive functioning response to the DP, they will perform better on the ERT than their Western counterparts. Yet empirical results show that despite the striking advan-tage that Chinese toddlers have in executive functioning skills – for example, on average, 42-month old Chinese children scored on par with 48-month old North American preschoolers (Sabbagh et al. 2006) – there was no significant difference in

---

[5] Importantly, the Dynamical Systems Theory can also provide nonrepresentational explanations of internal brain processes (see Freeman 1975; Chemero 2009). Note also that Occam's razor may apply here: if a given phenomenon can be explained without the need of adding an extra ingredient – i.e. representations – one should adopt the most parsimonious option.

their ERT. This supports the hypothesis that executive maturation alone does not account for preschoolers' false belief development.[6]

Another important concern regards the use of the term "metarepresentation". As S. Scott (2001) pointed out, there is danger of confusion – with serious consequences for the debate about the nature of higher-level cognition – between two distinct notions of "metarepresentation", as defined by philosophers (Dennett 1998) and by psychologists dealing with the question of autistic disorders (Leslie 1991; Baron-Cohen 1995). According to Dennett (1998), representations are themselves objects in the world, and therefore potential objects of (second-order or meta-) representations. Call this metarepresentation$_1$. For example, drawing a cat on a piece of paper is a type of non-mental representation, which is represented in the mind of the person viewing it. The mental representation is of the drawing, but since the drawing is itself a representation, the viewer has a (mental) metarepresentation of whatever it is that the drawing represents, namely a cat.[7] By contrast, Leslie uses the term "metarepresentation" to mean "(e.g., in the case of understanding pretence-in-others) an internal representation of an epistemic relation (PRETEND) between a person, a real situation and an imaginary situation (represented opaquely)…"(Leslie 1991:73) Call this metarepresentation$_2$. This definition does not sound at all like the definition of metarepresentation$_1$ as second-order representation pursued above. There is nothing metarepresentational in the sense of "higher order representation" in Leslie's formulation of the semantics of psychological predicates. Building on this distinction, S. Scott insightfully argues that a representation can contain other representations without being a metarepresentation$_1$.

Consider (P):
(P) The child BELIEVES that Sally BELIEVES that the marble is in the basket.

In what follows, I shall argue that although (P) is a straight-up second-order belief, this does not necessarily involve second-order representation, or metarepresentation$_1$ in Dennett's sense. Against De Bruin and Kästner (2012), I hold that there are no additional second-order "metarepresentational" skills involved in ERT as compared with VOE trials. Much of what I have to say in this section parallels arguments from Scott (2001) with which I am in close agreement. Scott convincingly argued that second-order beliefs do not necessarily require metarepresentations$_1$. It is only necessary to have the ability to represent first order beliefs in order to have second-order beliefs (Scott 2001: 940).

Take the following example of a first order belief:

(1)   Melissa BELIEVES that her dog is dead.

---

[6] The same findings have been replicated with Korean children (Oh and Lewis 2008). Moreover, developmental cognitive neuroscience studies suggest that the neural bases of mentalizing abilities are clearly dissociable from those that deal with executive functioning skills (Sabbagh et al. 2009; Saxe and Wexler 2005; Sommer et al. 2007).

[7] Note that this is a hybrid form of metarepresentation – the drawing being "external" and not a mental representation. Of course, one interesting question is whether hybrid metarepresentation is the same sort of thing as purely mental metarepresentation. However, this need not concern us here. The important point is that metarepresentation$_1$ is cognitively demanding and humans are unique in their capacity to form this type of metarepresentations.

The crucial point here is that to simply hold a belief, Melissa need not be aware of her belief or to hold an explicit representation[8] of it. In other words, she need not think to herself: "I believe my dog is dead" or "It is I who believes that my dog is dead". At this level of interpretation, we can speak of animals having this kind of online implicit beliefs, although we may find uncomfortable the idea of dogs having implicit beliefs. Now, consider the following example of a second order belief:

(2)   Anne BELIEVES that Melissa BELIEVES that her dog is dead.

As Scott rightly points out, in order to get (2) Anne needs the representation of Melissa's dog, the predicate DEAD, and so on. What she does not need is a representation of Melissa's representation of her dog, the predicate DEAD, and so on. That is, she does not need a second-order representation of any of these things. She can get by with her own first-order representations. Given that neither Melissa nor Anne has any particular need of belief representation in order to be a believer, Anne's representation of Melissa's belief need not be second order. In addition, it would seem that what Anne also needs in order to get (2) is a "representation of Melissa's BELIEF". The question is: is there any metarepresentation$_1$ involved here? Indeed, one might object[9] that Melissa's belief state involves already implicit or sub-personal representational processing. Now, the distinction between explicit versus implicit or sub-personal mental representations is a complicated issue and need not concern us here. For present purposes, it is sufficient to insist on the idea that Anne's holding a representation of Melissa's first-order belief (regardless of the fact that the latter involves or not subpersonal representational processing in Melissa's mind) does not amount to a second-order metarepresenting$_1$ (in Anne's mind). But let us suppose for the sake of the argument that Anne holds a representation of Melissa's first-order implicit belief (B) which in turn involves a certain sub-personal representational processing (S) in Melissa's brain. Now, if (S) is an implicit, sub-personal representation (in Melissa's mind), then one consequence would be that in metarepresenting Melissa's belief (B) [which involves (S)], Anne is not explicitly aware of what she is metarepresenting. Indeed, given that one member of the double representational layer, namely (S), remains opaque to her, Anne is aware only of what she is explicitly representing, namely (B). Note that this is not a problem per se. Let us label this half-opaque metarepresenting, metarepresentation$_3$. If this is so, then I fail to see why metarepresenting$_3$ in this sense is supposed to be cognitively more demanding (for Anne) than mere representing. In contrast, recall that in Dennett's drawing example, the viewer is fully aware of the double representational layer: he forms a mental representation *of* a representation (drawing) *of* a cat and this makes his metarepresenting$_1$ a genuine second-order cognitive achievement. Hence, it is not clear that higher-order metarepresenting$_1$ is at work in the Sally/Anne scenario and this casts doubt on the idea that the ERTs require that infants not only represent but metarepresent.

To sum up, according to De Bruin and Kästner, ERTs involve a stronger form of decoupling (precisely because it involves metarepresentational skills and language

---

[8] Explicit representations are available to one's explicit awareness, whereas sub-personal or implicit representations are processed without one's explicit awareness.

[9] I am grateful to the anonymous reviewer who pressed clarification on this paragraph.

processing), hence explaining the Developmental Puzzle. Although I agree with De Bruin and Kästner in saying that (a) ERTs require decoupling, and that (b) the verbal interaction with the experimenter during ERT plays a crucial role in 3-year olds' failure to report false-belief understanding, there is still something missing in the picture. Indeed, I fail to see how (a) and (b) alone can solve the Developmental Puzzle, since, as the authors themselves have insisted, the decoupling is supposed to lead to an optimization of cognitive processing. Everybody agrees that strong decoupling is an important evolutionary advantage. But the mystery of the Developmental Puzzle stems from the opposite situation. In order to truly solve the DP, they need to answer the following question: why does stronger decoupling impair (at least in some cases) rather than improve the "mental gymnastics" of representational manipulation? In other words: why do weaker forms of decoupling do a better job in a complex task such as false-belief understanding?

As I hope has become clear throughout this section, although DEC leaves the Developmental Paradox unsolved, there are important aspects in De Bruin and Kästner's account that could help us to unveil the mystery. For example, one might argue that the "coupling /decoupling/ recoupling" loop serves cognition for better and for worse. That is, in certain circumstances, this dynamic interplay can alter rather than enhance the online processing. But in what circumstances exactly? And what are the most disturbing elements in the ERT scenarios? In the final section of this paper, I suggest an enactive response to these puzzles.

4 An Enactive Solution to the Puzzle

Let me begin with a short thought experiment. Imagine you are sitting in a soft, warm and comfortable place enjoying a piece of music. You are fully immersed in your auditory experience. After the music stops, suppose I debrief you by asking for a detailed verbal report about your recent experience. If my line of reasoning is correct, I predict that the more active and pressing our interactions are, the poorer your verbal-report performance will be.[10] This is because instead of leisurely building decoupled representations of your recently experienced scenario, your priority now is to process and regulate the online social interaction with me, the experimenter. Indeed, it is crucial for our survival as human animals to successfully deal with moment-to-moment processing first, before safely withdrawing and manipulating offline (decoupled) information.

With this is mind, let us return to the Developmental Puzzle. Recall that visual perception plays a crucial role in processing the implicit false-belief understanding as opposed to the explicit/verbal-report condition. But why is perception, in some cases, "smarter" than explicit and verbalized thinking? In order to deal with this question, it is perhaps helpful to proceed in two steps. Part (A) suggests that preverbal infants' success during VOE trials involves a pragmatic, action-oriented form of understanding based on direct perceptual processing of others' intentional attitudes rather than on

---

[10] A recent study (Rubio-Fernández 2013) illustrated that perspective tracking is a continuous process that can be easily disrupted in adults by a subtle visual manipulation in both indirect and direct false-belief tasks. I am grateful to the anonymous referee who directed me to this reference.

false-belief computing. Part (B) accounts for 3-year olds' failure during ERT and argues that the verbal-report interaction between the child and the experimenter elicits representational (and not metarepresentational) decoupling and may fall apart under pressure. In what follows, I shall address each part in turn.

(A) Unlike De Bruin and Kästner, I reject the idea that basic forms of mentality are representational and that during VOE scenarios, infants must rely on internal representations of visual information that is available to the other agent but not available to them. Rather, infants understand others' intentional attitudes as currently and readily available (i.e. directly observable) in the environment. To support this claim, I appeal to empirical findings illustrating that (i) infants' ability to understand other minds is rooted in their capacity to actively engage in interactive scenarios; and (ii) knowledge of visual perspective-taking is a relatively late cognitive achievement that is derivative from social[11] perspective-taking. Then I argue that these findings might raise a puzzle for De Bruin & Kästner's representationalist decoupling strategy. The overarching aim is to illustrate that: (a) nonrepresentational embodied practices might be at work and explain basic forms of intersubjective engagement; and (b) people's perceptions depend critically on their interactions with the environment rather than on internal information processing based on external visual observations (Brockmole et al. 2013).

Consistent with a burgeoning literature suggesting a common basis for both the production and perception of action, evidence has been mounting to illustrate that infants' understanding of others is more robust within interactive contexts (Király 2009; Moll et al. 2008). In other words, the more engaged the interactions infants/agents are the more robust the infants' understanding of others becomes.[12] Before their first birthday, preverbal human infants not only understand what goals or preferences others have (Gergely et al. 1995; Repacholi and Gopnik 1997); they also track what others know, in the sense of what they have and have not experienced in the immediate past. Indeed, several studies suggest that visual perspective-taking does not precede ontogenetically but follows social perspective-taking (Moll and Kadipasaoglu 2013). Children first learn to discern or establish reference in situations that are not defined by differences in how self and other perceive agents and objects visually but by differences in their shared experiential backgrounds, i.e. in what they did, witnessed or heard. For example, Moll and Tomasello (2007) tested the child's abilities to recall an adults' knowledge of what she has experienced in three conditions: (1) the child and adult together interacted with a toy; (2) the infant handled the toy with another experimenter, while the adult watched (and the infant was alerted to this several times); (3) the adult handled a toy alone, while the infant watched. As Wilby (2012) pointed out, one might describe the difference in evidence that is available to the infant as follows:

(1)   X is aware that [I am aware that [X is aware that [p]]].

---

[11] Social perspective-taking is a set of manifold abilities of infants which helps them to establish reference against the background of prior social interactions.

[12] For example, Sommerville et al. (2005) demonstrated that 3-month olds focus on the relation between an agent and her goal if they reached for (and not merely looked at) a toy before observing another actor grasping it. Also 10-month olds who received active training in pulling a cane to retrieve a toy subsequently registered another person's cane-pulling action as goal-directed behavior, while infants who underwent mere observational training were unable to do this. Importantly, goal-relatedness is differently perceived by infants in social versus physical event configurations (Woodward et al. 2001).

(2)   X is aware that [I am aware that [p]].
(3)   X is aware that p.

Now, if we apply De Bruin and Kästner's "degrees of decoupling" explanatory strategy in this specific case, then one could expect that infants would find the first condition (1) the hardest, since it involves several embedded "layers" of decoupling. Yet, the evidence suggests the complete opposite. Hence, it is not clear that crediting infants with an implicit representational decoupling ability is the best strategy here. Of course, this is not a knockdown argument against the representationalist approach. But endorsing a radically enactivist account avoids this problem and could potentially explain basic forms of joint attending, those of the non-linguistically-mediated online sort, by stressing the crucial role of embodied interactions (Gallagher and Varga 2013). Indeed, the subject's understanding of a scenario corresponds to the subject's potential for skilful interaction with worldly objects and social agents involved in that scenario. Furthermore, infants' own actions inform their perceptions of goal-directed behaviour (Falck-Ytter et al. 2006) and thereby contribute to their ability to view others as goal-directed agents. Hence, this (re)-enacted relationship with the world "suffused with practical understanding, is what perceptual experience essentially is" (Ward 2012: 738, original emphasis).

(B) The second part of my argumentation concerns the explicit question (P) the child faces during ERT trials:

(P) Where will Sally look for her marble?

Note that there are at least two implicit components that situate this embodied interaction within a specific context and thereby play a crucial role during ERT, namely the verbal-report interaction component (Q_1) and the time–pressure component (Q_2):

(Q_1)   You tell me (P).
(Q_2)   You tell me now (P).

Importantly, during VOE trials, infants are "left alone" to watch/process a scene at their leisure: there is no offline-based task to cope with, and no additional verbally interactive ingredients to disturb them from enjoying/understanding the story happening live, so to speak. In contrast, dealing with an explicit verbal-report demand places a supplementary pressure on the child's shoulders. Now, it is crucial for my argumentation here to distinguish between verbal expressions of (P) and verbal interactions *about* (P). Indeed, someone might object[13] that already in a study by Clements and Perner (1994) when the experimenter wondered aloud to himself, 'I wonder where will Sally look for her marble?' children immediately looked at the right place as a rule. But compare (Q_1) with (Q_3):

(Q_1)   You tell me (P).
(Q_3)   (I wonder (P)).

---

[13] I am indebted to comments from an anonymous referee for pressing the clarification in this paragraph.

In both cases, the experimenter verbally expresses (P).[14] But whilst in (Q$_3$) the child is passively (leisurely) witnessing a self-addressed monologue, in (Q$_1$) he is actively invited to engage into an interactive dialogue. Hence, the disruption is not caused by the mere verbal expression of (P), i.e. saying (P) aloud, but by the verbal interaction *about* (P) between two relating agents.[15] And this is where the representational decoupling steps into the picture, since in (Q$_1$) the child has to take (P) offline. Indeed, providing a verbal report to the experimenter's question during the ERTs becomes a genuine offline representation-based decoupling task: the child has to regulate the intricate connection between her own takes on the world (coupling), Sally and Anne's remote doings (offline decoupling) and experimenter's demand here and now (online recoupling), and he has to do it fast, since the interrogator expects an immediate response. There is simply no time to build up a full-blown mental model of the scene, from which to derive a detached response, since during the ERT trials infants are not allowed to step back and leisurely process the scene in order to give a satisfactory answer. Indeed, it has been argued (Wilson 2002) that whilst online sensorimotor-based coupling is designed to deal with real time/ run-time tasks, representational decoupling as the one involved in verbal reports is time-costly and may fall apart under time–pressure since it necessitates the fulfilment of specific conditions: for this very reason, even adults do not employ it all the time, by default and in all situations.[16] Indeed, one needs to secure the task-relevant online scene first in order to safely withdraw and leisurely engage in metarepresenting, hypothetical planning, counterfactual reasoning, etc. However, if one lifts the demands of time pressure, some of the true power of human cognition becomes evident. To return to our thought experiment: if I leave you alone after your musical experience, with the command of reporting your experience, you might actually be able to create a masterpiece. But you will definitely need to process the scenario at your leisure.

To sum up: I argued here that when we mentally decouple from an event E, we are already a step after experiencing E via our skilful interaction with worldly objects and social agents. To put it in a slogan: decoupling processing supervenes on coupling processing. This means that changes at the online coupling/recoupling level will trigger changes at the decoupling processing level (if there is an offline take activated). Hence, the primacy of the former over the latter might explain why 3-year olds give priority to the online social interaction task and leave the decoupling representation-based task behind.

Recall that the cognitive gap challenge is supposed to prompt enactivism defenders to tell a story about how low-level sensorimotor practices transform into higher-order representational skills. Indeed, one might object[17] that the only way to avoid the gap is

---

[14] Note that the verbal expression of (P) might be absent altogether, as in the "helping paradigm" trials (Buttelman et al. 2009). This corroborates the idea that the disrupting ingredient in ERT is not the mere verbal component, but the verbally interactive one.

[15] A recent study (Rubio-Fernández and Geurts 2013) illustrates that nonverbal versions of the ERT allow infants to keep track of a protagonist's perspective over a course of events, whereas verbal designs tend to disrupt the perspective tracking process.

[16] For example, Keysar et al. (2003) have shown that adults' first rapid guesses about the meaning of words (using eye-tracking detecting measures) are based more on empirical generalizations (e.g. how a speaker has used a word in the past) rather than on complicated inferences about speakers' beliefs and intentions. This does not imply that adults are unable to make such sophisticated inferences. However, these operations take time.

[17] I am grateful to the anonymous referee who pressed clarification on this point.

to endorse eliminativism and to deny the existence and importance of contentful and representation-based forms of cognition. Leaving aside the fact that several theorists have convincingly illustrated that the sensorimotor system actually has the right kind of structure to characterize both sensorimotor and abstract concepts (Barsalou 1999; Gallese and Lakoff 2005; Pecher et al. 2004; Hauk et al. 2004), it seems to me that there are at least two additional strategies that REC proponents could endorse without falling into eliminativism. The easy way consists in sending the ball into the opposite court: if everybody agrees that the preverbal infant's understanding is grounded in perceptual activity, then cognitivists and reconciliatory defenders should come up with a positive account of how exactly the implicit decoupling, supposed to be found already at work in preverbal human infants, is "grounded in enaction". That is, they have to provide a story about how the alleged decoupled (albeit partly) representations are implemented in the brain's systems.

Another strategy, that I am inclined to endorse here, consists in dissolving the gap by arguing that the problem might arise as a side effect of a coupling/decoupling/recoupling investigatory interplay. Indeed, when we start to investigate cognition (both theoretically and empirically), we inevitably address it through a high-level cognitive lens which might distort our take on the original underlying processing which we find already there, at work, with no gaps in it.

5 Conclusion

I have argued here that the reconciliatory DEC strategy (De Bruin and Kästner 2012) leaves the Developmental Puzzle unsolved. More specifically, I have argued that their commitment to the (rich) representationalist thesis according to which preverbal human infants run offline/decoupled processing (albeit partial) prevents De Bruin and Kästner from solving the paradox. While I agree with them in their claim that failures in the coupling/decoupling /recoupling dynamic interplay are at work during ERT, I disagree with them when they attribute these failures to stronger (metarepresentational) decoupling demands. Rather, I suggested, infants fail because the ERT trials involve representational decoupling *tout court* and what is more, under pressure.

References

Agre, P., and D. Chapman. 1987. *Pengi: An Implementation of a Theory of Activity. In Proceedings of the Sixth National Conference on Artificial Intelligence, 268–272*. Menlo Park: American Association for Artificial Intelligence.

Apperly, I., and S. Butterfill. 2009. Do humans have two systems to track beliefs and belief-like states? *Psychological Review* 116: 953–970.

Baillargeon, R., R.M. Scott, and H. Zijing. 2010. False-belief understanding in infants. *Trends in Cognitive Science* 14(3): 110–118.

Baron-Cohen, S. (1995). Mindblindness: an essay on autism and theory of mind. MIT Press/Bradford Books.

Barsalou, L.W. 1999. Perceptual symbol systems. *Behavioral and Brain Sciences* 22: 577–660.

Bermúdez, J. 2003. *Thinking Without Words*. New York: Oxford University Press.

Brockmole, J.R., C.C. Davoli, R.A. Abrams, and J.K. Witt. 2013. The world within reach: Effects of hand posture and tool use on visual cognition. *Current Directions in Psychological Science* 22(1): 38–44.

Buttelman, D., M. Carpenter, and M. Tomasello. 2009. Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition* 112: 337–342.

Carlson, S.M., and L.J. Moses. 2001. Individual differences in inhibitory control and children's theory of mind. *Child Development* 72: 1032–1053.

Carlson, S., L. Moses, and C. Breton. 2002. How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development* 11: 73–92.

Chang, F.-M., J.R. Kidd, K.J. Kivak, A.J. Pakstis, and K.K. Kidd. 1996. The world-wide distribution of allele frequencies at the human dopamine D4 receptor locus. *Human Genetics* 98: 91–101.

Chemero, A. 2009. *Radical Embodied Cognition*. Cambridge: MIT Press.

Chen, X., P.D. Hastings, K.H. Rubin, H. Chen, G. Cen, and S.L. Stewart. 1998. Childrearing attitudes and behavioral inhibition in Chinese and Canadian toddlers: A crosscultural study. *Developmental Psychology* 34: 677–686.

Churchland, P.S. 2002. *Brain-Wise*. Cambridge: MIT Press.

Clark, A. 1997. *Being there: putting mind, body, and world together again*. Cambridge: MIT Press.

Clements, W.A., and J. Perner. 1994. Implicit understanding of belief. *Cognitive Development* 9: 377–395.

Cole, K., and P. Mitchell. 2000. Siblings in the development of executive control and a theory-of-mind. *British Journal of Developmental Psychology* 18: 279–295.

De Bruin, L.C., and L. Kästner. 2012. Dynamic Embodied Cognition. *Phenomenology and the Cognitive Sciences* 11(4): 541–563.

Dennett, D. (1998). Making tools for thinking. In D. Sperber (Ed.) (2000). Metarepresentations. New York: Oxford University Press.

Dretske, F. 1995. *Naturalizing the mind*. Cambridge: MIT Press.

Falck-Ytter, T., G. Gredebäck, and von Hofsten. 2006. Infants predict other peoples' action goals. *Nature Neuroscience* 9: 878–879.

Freeman, W. 1975. *Mass Action in the Nervous System*. New York: Academic.

Gallagher, S. 2005. *How the Body Shapes the Mind*. Oxford: Oxford University Press.

Gallagher, S. and Varga S. (2013). Social Constraints on the Direct Perception of Emotions and Intentions. Topoi:1–15 (2013).

Gallese, V., and G. Lakoff. 2005. The brain's concepts: The role of the sensory-motor system in reason and language. *Cognitive Neuropsychology* 22: 455–479.

Gergely, G., Z. Nadasdy, G. Csibra, and S. Bıró. 1995. Taking the intentional stance at 12 months of age. *Cognition* 56: 165–193.

Grush, R. (1997). Yet another design for a brain? Review of Port and van Gelder (Eds.), Mind as motion. *Philosophical Psychology*, 233–242.

Hala, S., S. Hug, and H. Henderson. 2003. Executive functioning and false-belief understanding in preschool children: two tasks are harder than one. *Journal of Cognition and Development* 4: 275–298.

Hauk, O., I. Johnsrude, and F. Pulvermüller. 2004. Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41: 301–307.

Ho, D.Y.F. 1994. Cognitive socialization in Confucian heritage cultures. In *Cross-cultural roots of minority development*, ed. P.M. Greenfield and R.R. Cocking, 285–313. Hillsdale: Erlbaum.

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

Hutto, D., and E. Myin. 2013. *Radicalizing Enactivism*. Cambridge: MIT Press.

Keysar, B., S. Lin, and D.J. Barr. 2003. Limits on theory of mind use in adults. *Cognition* 89: 25–41.

Király, I. 2009. Memories for events in infants: Goal-relevant action coding. In *Social Cognition: Development, Neuroscience and Autism*, ed. T. Striano and V. Reid. Oxford: Wiley-Blackwell.

Leslie, A.M. 1991. Precursors to a theory of mind. In *Natural Theories of Mind: Evolution, Development, and Simulation of Everyday Mindreading*, ed. Whiten Andrew, 63–78. Oxford: Blackwell.

Leslie, A.M., T.P. German, and P. Polizzi. 2005. Belief–desire reasoning as a process of selection. *Cognitive Psychology* 50: 45–85.

Menary R. (2007). Cognitive Integration: Mind and Cognition Unbounded. Palgrave.

Millikan, R. 1995. Pushmi-Pullyu Representations. *Philosophical Perspectives* 9: 185–200.

Moll, H., and Kadipasaoglu, D. (2013). The primacy of social over visual perspective-taking. *Frontiers in Human Neuroscience, 7*(558). doi: 10.3389/fnhum2013.00558.

Moll, H., and M. Tomasello. 2007. How 14- and 18-month-olds know what others have experienced. *Developmental Psychology* 43(2): 309–317.

Moll, H., N. Richter, M. Carpenter, and M. Tomasello. 2008. Fourteen-month-olds know what 'we' have shared in a special way. *Infancy* 13(1): 90–101.

Oh, S., and C. Lewis. 2008. Korean preschoolers' advanced inhibitory control and its relation to other executive skills and mental state understanding. *Child Development* 79: 80–99.

Onishi, K.H., and R. Baillargeon. 2005. Do 15-Month-Old Infants Understand False Beliefs? *Science* 308(8): 255–258.

Pecher, D., and R. Zwaan. 2005. The Role of Perception and Action in Memory, Language, and Thinking. In *Grounding cognition: The role of perception and action in memory, language, and thought*, ed. D. Pecher and R. Zwaan. New York: Cambridge University Press.

Pecher, D., R. Zeelenberg, and L.W. Barsalou. 2004. Sensorimotor simulations underlie conceptual representations: Modality-specific effects of prior activation. *Psychonomic Bulletin & Review* 11: 164–167.

Perner, J., and B. Lang. 1999. Development of theory of mind and executive control. *Trends in Cognitive Sciences* 3(9): 337–344.

Perner, J., and T. Ruffman. 2005. Infants' insight into the mind: How deep? *Science* 308: 214–216.

Perner, J., B. Lang, and D. Kloo. 2002. Theory of Mind and Self Control: More than a common problem of inhibition. *Child Development* 73: 752–767.

Repacholi, B.M., and A. Gopnik. 1997. Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology* 33: 12–21.

Ridderinkhof, K.R., M. Ullsperger, E.A. Crone, and S. Nieuwenhuis. 2004. The role of the medial frontal cortex in cognitive control. *Science* 306: 443–447.

Rubio-Fernández, P. 2013. Perspective tracking in progress: Do not disturb. *Cognition* 129(2): 264–272.

Rubio-Fernández, P., and B. Geurts. 2013. How to pass the false-belief task before your fourth birthday. Psychological Science. *Jan 1;* 24(1): 27–33.

Sabbagh, M.A., F. Xu, S.M. Carlson, L.J. Moses, and K. Lee. 2006. Executive functioning and theory-of-mind in preschool children from Beijing, China: Comparisons with U.S. preschoolers. *Psychological Science* 17: 74–81.

Sabbagh, M.A., L.C. Bowman, L.E. Evraire, and J.M.B. Ito. 2009. Neurodevelopmental correlates of theory of mind in preschool children. *Child Development* 80: 1147–1162.

Saxe, R., and L.J. Powell. 2006. It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychological Science* 17: 692–699.

Saxe, R., and A. Wexler. 2005. Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia* 43: 1391–1399.

Scott, S. 2001. Metarepresentations in Philosophy and Psychology. In *Proceedings of the twenty-third annual conference of the cognitive science society, University of Edinburgh*, ed. J. Moore and K. Stenning. London: LEA Publishers.

Scott, R.M., R. Baillargeon, H.-J. Song, and A. Leslie. 2010. Attributing false beliefs about non-obvious properties at 18-months. *Cognitive Psychology* 61(4): 366–395.

Sodian, B., C. Thoermer, and N. Dietrich. 2006. Two- to four-year-old children's differentiation of knowing and guessing in a non-verbal task. *European Journal of Developmental Psychology* 3: 222–237.

Sommer, M., K. Dçhnel, B. Sodian, J. Meinhardt, C. Thoermer, and G. Hajak. 2007. Neural correlates of true and false belief reasoning. *NeuroImage* 35: 1378–1384.

Sommerville, J. A. and Woodward, A., (2010). The Link Between Action Production and Action Processing in Infancy. In Grammont, F., Legrand D., Livet, P. (eds.) Naturalizing intention in action, Cambridge, Mass.: MIT Press.

Sommerville, J.A., A. Woodward, and A. Needham. 2005. Action experience alters 3-month-old infants' perception of others' actions. *Cognition* 96: B1–B11.

Spaulding, S. 2010. Embodied cognition and mindreading. *Mind & Language* 25(1): 119–140.

Surian, L., S. Caldi, and D. Sperber. 2007. Attribution of beliefs to 13-month-old infants. *Psychological Science* 18: 580–586.

Tobin, J.J., D.Y.H. Wu, and D.H. Davidson. 1989. *Preschool in three cultures: Japan, China and the United States*. New Haven: Yale University Press.

Ward, D. 2012. Enjoying the Spread. *Conscious Externalism Reconsidered. Mind.* 121(483): 731–751.

Wellman, H.M., D. Cross, and J. Watson. 2001. Meta-analysis of Theory of Mind Development: The Truth about False-Belief. *Child Development* 72(3): 655–684.

Wilby, M. (2012), Embodying the False-Belief Tasks. Phenomenology and the cognitive sciences, special issue on 'debates on embodied mindreading' (ed. Shannon Spaulding) December 2012, Volume 11, pp. 519–540.

Wilson, M. 2002. Six views of embodied cognition. *Psychonomic Bulletin and Review* 12(9): 625–636.

Wimmer, H., and J. Perner. 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13: 103–128.

Woodward, A.L., J.A. Sommerville, and J.J. Guajardo. 2001. How infants make sense of intentional action. In *Intentions and intentionality: Foundations of social cognition*, ed. B. Malle, L. Moses, and D. Baldwin. Cambridge: MIT Press.