

Enactive and Behavioral Abstraction Accounts of Social Understanding in Chimpanzees, Infants, and Adults

Shaun Gallagher · Daniel J. Povinelli

Published online: 2 March 2012
© Springer Science+Business Media B.V. 2012

Abstract We argue against theory-of-mind interpretation of recent false-belief experiments with young infants and explore two other interpretations: enactive and behavioral abstraction approaches. We then discuss the differences between these alternatives.

Over the past 30 years, developmental psychologists have unveiled surprising competences in very young infants, with equally surprising claims about the cognitive systems supporting these competences. Recent experiments with young infants following the “false-belief” paradigm have reinforced this idea in the realm of social cognition (Onishi and Baillargeon 2005; Senju et al. 2011). One question these experiments raise concerns the precise nature of the infant’s ability to understand the behaviors and mental states of others. Comparative psychologists have made similar claims about the social cognitive abilities of chimpanzees and other animals. Are these claims warranted by the evidence? To different degrees, both of us have argued that they are not.

The two most developed and defended answers have been cast in terms of a behavioral approach (e.g., behavioral abstraction, cf. Povinelli and Vonk (2003, 2004); Penn and Povinelli (2007) or behavioral rules, cf. Perner and Ruffman (2005)) and a theory-of-mind (ToM) approach (either theory theory, cf. Onishi and Baillargeon (2005); Carruthers (2009) or simulation theory, cf. Herschbach (2007)). In this paper we evaluate these theories, as well as an alternative account in terms of enactive perception, in an effort to specify abilities that are basic for both chimpanzees and young infants, and remain basic for adult social cognition. In the first

S. Gallagher (✉)
Department of Philosophy, University of Memphis, Memphis, TN, USA
e-mail: s.gallagher@memphis.edu

S. Gallagher
School of Humanities, University of Hertfordshire, Hatfield, UK

D. J. Povinelli
Cognitive Evolution Group, University of Louisiana, Lafayette, LA, USA

section, Shaun Gallagher reviews recent evidence on false belief experiments in infants and rejects metarepresentational, simulationist and rule-based accounts in favor of an enactive interpretation. In the second section, Daniel Povinelli analyzes results in the chimpanzee in terms of behavioral understanding rather than mental understanding. In the third section, we evaluate to what extent there may be continuity between infants, chimps and human adults. Finally, in the last section, we discuss, in a conversational mode, the similarities and differences between our two views.

1 False Belief Experiments with Infants: An Enactive Account

We begin with the false-belief experiments conducted with very young (15- and 13-month-old) infants (Onishi and Baillargeon 2005; Surian et al. 2007). Carruthers (2009, p. 166), for example, citing just these studies, suggests that infants already have metarepresentational “mindreading” abilities. This is one possible interpretation of the empirical evidence – but a controversial one that is currently under debate and that suggests a closer look at the evidence.

Many of the experiments supporting early mental state (in this case, belief) reasoning in infants involve the following pattern. The infant observes an agent place a toy in a box (A). The agent leaves the room. The infant then sees the toy moved to a different location (B). When the agent returns the infant looks to A as the location where she expects the agent to look for the toy – anticipatory looking (Southgate et al. 2007) – and if the agent goes directly to B to look for the toy the infant exhibits a violation of expectation (VOE) that signals surprise (Onishi and Baillargeon 2005; Song et al. 2008; Surian et al. 2007).

First, we note that the experiments are consistently designed and presented as experiments that test false belief comprehension. That is, they are described in terms of a test for the infant’s ToM – their comprehension of what the other believes. From the perspective of ToM approaches, however, many have found these experimental results surprising because the consensus had been that infants this young (15 months and younger) did not yet have a concept of belief, and were certainly not able to represent (or engage in the kind of metarepresentational process necessary to grasp) false belief. Three-year-old children consistently fail standard (verbal-based) false-belief and related meta-representational tests (e.g., see meta-analysis by Wellman et al. 2001) so it is surprising to find 13- and 15- month-old infants passing such tests, despite differences in experimental design. Viewing the experiments from the perspective of ToM, and certainly deploying their own ToM, the experimenters are naturally led to say that very young infants are capable of recognizing false beliefs:

In these tasks, children’s understanding of an agent’s false belief is inferred from behaviors they spontaneously produce as they observe a scene unfold (just as adults watching a movie might spontaneously produce responses that reveal their understanding of the characters’ mental states). (Baillargeon et al. 2010, p. 110)

There is no doubt that we have a clear case of metarepresentational mindreading here, at least on the part of the scientists. The question is whether the infant’s understanding is also based on inference from evidence that leads to recognition of false belief.

In each case we can say that the infant expects a certain action, an expectation formed by familiarity with a situation in which the agent sees or does not see something. The infant knows, for example, that the agent has not seen where the toy has been placed, anticipates that the agent will look one place, but is surprised that the agent looks in a different location. Baillargeon et al. (2010) conclude that the infant not only infers that the agent's mental state consists of a false belief, but that the child can reason about a complex set of mental states, including dispositional preferences, intended goals, knowledge about the situation, inferences, and false beliefs.

One might equally claim that the infant's expectations are shaped by the infant's taking into account that the agent's actions are informed by her (the agent's) familiarity with the situation, and specifically by what that agent has been in a position to see or not see. The agent has not been in a position to see the toy's switch from location A to location B. If the agent's actions are guided by what she has been in a position to see, then one would expect her to look in A and is surprised that she looks in B. Here's the scenario.

- (1) Agent puts toy in A, or sees toy put in A.
- (2) Infant sees (1) – that is, the infant sees that the agent is in a position to see where the toy is.
- (3) The toy is shifted from A to B, but the agent is not in a position to see this.
- (4) Infant sees (3) – that is, the infant sees that the agent is not in a position to see the shift.
- (5) TT hypothesis: Infant infers that agent has false belief about location of toy.
- (6) When agent looks in B, there is a VOE.

One could contend that TT introduces an extra step (5), where the infant has to make an inference to a mental state.

Simulation theory (ST), in its classical formulation, also posits extra steps, namely the formation of a pretend belief and the projection of that belief to the other's mind. In its more recent neuronal formulation, this projection is automatic and accomplished by activation of mirror neurons (MNs). Herschbach (2007), adopting Dokic's (2002) idea that simulation may account for the infant's ability in this regard, suggests that the performance of young infants on these false-belief tests are 'online' (and implicit) rather than metarepresentational. According to Herschbach, simulation involves "the information about the other's beliefs gained from pretending to have those beliefs (where 'pretending' is not necessarily a conscious or person-level notion)" (2007, p. 15). What precisely a sub-personal pretending is, however, is not clear. If one equates simulation with activation of MNs, as, for example, Gallese, Goldman, and others do, and at the same time, if one holds that the mirror system is "neutral" with regard to who the agent is, since it is activated in both self-action and the observation of another's action, then this sub-personal system cannot involve pretense, something which necessarily involves a self-other distinction. There would be no representation of "*my own* motor action *as if* it were *yours*" if indeed there is no *you* or *I* represented in MNs. Even if there is differentiation between *you* and *I* implicit in the firing rates of MNs, pretense requires even more than that basic differentiation (see Gallagher 2008).

Such worries have motivated Goldman to defend a more minimal and generic concept of simulation based on what can be termed the 'matching

hypothesis'. Simulation exists in a minimal generic sense if one system matches the other (Goldman and Sripada 2005; Goldman 2006). A number of empirical studies challenge this idea, however. Catmur et al. (2007), for example, show that, with learning, incongruent (non-matching) actions can activate MNs with no loss of understanding on the part of the observer or actor (also see Dinstein et al. 2008). Csibra (2005) suggests that only about one-third of MNs show a one-to-one congruence, motivating the suggestion that MNs may be involved in preparing a complementary action in response to an observed action (Newman-Norlund et al. 2007, p. 55). Accordingly, even if MNs are operational in the infant's brain, they are not subtending a form of simulation that involves either pretense or matching.

In contrast to both TT and ST, a more behavioral approach suggests that the infant expects the agent's action to be guided by whether the agent has been in a position to see or not see something. In this case the infant already has enough information (2 and 4) to explain the VOE at (6). The infant has a pragmatic grasp of how perception and action are connected and does not have to infer or simulate anything about mental states or beliefs in order to understand the agent. Thus the behavioral rules view proposed by Ruffman and Perner (2005) appeals to the infant's grasp of behavioral rules (e.g. 'people look for objects where they last saw them') gained via statistical learning.

Baillargeon et al. (2010) suggest that the behavioral rules explanation fails because of the large number of rules that would be needed in a variety of situations involving false beliefs. On the one hand, if a variety of different behaviors target different characteristics of the objects (e.g., the toy) involved – identity versus properties versus number, etc. – this could imply a multiplication of behavioral rules. It would be unlikely that infants this young would come equipped with, or be able to learn so many rules and use them so readily. Such rules may even be more complex in circumstances where the infant is guided by knowledge gained in different ways. On the other hand, advocates of behavioral rules might respond that it is not clear why infants should not be able to apply a lesser number of more general, sensory and modality-independent, perception and action principles to different circumstances, or apply them flexibly to information gained via different sensory or motor modalities, especially since infants spend their entire first year interacting with others and begin to engage in joint attention and joint actions starting around 9–12 months. Träuble et al. (2010), for example, suggest that if such rules are formulated more flexibly or more generally, the behavioral rules view cannot be ruled out, since: "infants might well use a rule whereupon people search for objects according to their perceptual access in a more general sense, including various forms like visual, auditory, and tactile access" (pp. 442–443; but also see Stack and Lewis 2008 for further criticism).

An alternative to either ToM or behavioral approaches is based on the concept of enactive perception. This approach rejects TT and ST, and it avoids some of the problems of the behavioral rules approach, especially the idea that the infant makes an inference based on a rule. It also involves a different, non-simulationist interpretation of the mirror system and by doing so it stays close to the level of embodied behavior. This enactive approach to social cognition is part of a larger phenomenological theory sometimes referred to as 'interaction theory' (Gallagher 2005; Gallagher and Zahavi 2008; also De Jaegher et al. 2010; de Bruin et al. 2011). If perception is an enactive

process (Hurley 1998; Noë 2004; O'Regan and Noë 2001; Varela et al. 1991) then it is more appropriate to think of mirror resonance processes as part of the structure of the perceptual process when it is a perception of another person's actions in a way that anticipates the possible actions I would take if I were to respond and engage with that agent. I see the other's action as something I can respond to – an intersubjective affordance – but this does not mean that I see it “as if” I were doing it, or that my motor system reacts as if it were engaged in the same action.

Without appealing to behavioral rules, the model of enactive perception suggests that infants perceive the world, and others, in terms of perceptual-action contingencies. Infants normally spend their entire first year interacting with others and at around 9–12 months begin to engage in joint attention and joint actions. Primary intersubjective abilities that characterize the infant's behavior from birth onward and allow them to interact with others, understanding their movements, facial expressions, gestures, etc. as meaningful (Trevarthen 1979; Reddy 2008), are supplemented by secondary intersubjective abilities that involve joint attention and reference to pragmatic contexts of action starting around age one year (Trevarthen and Hubley 1978). On the enactive view, infants understand others pragmatically in terms of how they can interact with them, or in terms of the infant's engagement in what the other is doing or expressing or feeling. Perception, in this sense, is not only *for action* (as Noë (2004), for example, makes clear), but also for *interaction*; as such it involves a pragmatic grasping of the other's behaviors and intentions in terms of social affordances. As Merleau-Ponty (1964, p. 119) puts it, “the other's intentions somehow play across my body” as a set of possibilities for my action. I see the other's action as aimed at the world in ways that offer affordances for interacting. This applies even in cases where I am simply *observing* rather than actually interacting with others. The way that agents are involved in the world, which I see, can influence my expectations of how I might act in the world and interact with them.

In other words, seeing what the other is doing is framed in terms of the possible actions I would take if I were to engage with that agent. This, however, does not require making inferences to the agent's mental states. This can be seen in some of the more interactive experimental designs used to study infants' social cognition. In a study by Buttelmann et al. (2009) 18-month-olds try to help an agent retrieve a toy while taking into account the fact that the agent doesn't know about a switched location (the false belief situation). In that situation, when the agent focuses on the wrong container (the original location, A), the infant is ready to lead him to the correct box (B), but not in the situation when the agent does know about the switch, i.e., the true belief situation, and still goes to A. In the latter case the infant goes to assist the agent at A. In this study when the agent goes to Box A, the infant sees exactly the same thing in the case of true belief (when the agent knows there has been a shift from A to B) as in the case of false belief (when the agent does not know about the shift). Again, the fact that the infant knows either that the agent has seen the switch or not, plus the agent's situated behavior with respect to A (e.g., moving to A and attempting to open it), is enough to specify the difference in the agent's intention. For the infant, there is no need to infer anything about true or false beliefs; the agent's behavior signals a difference in affordance, i.e., a difference in how the infant can act, and thereby interact with the agent. The infant does not have to make inferences to mental states since

all of the information needed to understand the other and to interact is already available in what the infant has seen of the situation.

Similar considerations hold for a study by Southgate et al. (2010), where the agent hides two toys in separate boxes, and then leaves. Infants then watch as another person switches the contents of the two boxes. When the agent returns she (the agent) points to one of the boxes (A), announcing that the toy hidden inside is a 'sefo'. When the infants are then asked to retrieve the 'sefo' most of them approach the other box (B), indicating that they must have understood that the agent intended to name the toy that was now in B, unaware of the toy's changed location. The agent's intention is not conceived in terms of a mental state; her intention is apparent in her performance, and her performance is part of what constitutes that intention. The infant sees the agent's original action and sees the switch that the agent does not see. There is interaction when the agent communicates in this situation and when the infant is invited to act. The infant does not have to engage in mindreading since all of this information is available in the behavioral situation, and is sufficient to inform the infant's action.

Generally speaking, the results of the false-belief experiments with very young infants suggest that the capacity for understanding social situations complicated by an agent's lack of information is closely intertwined with the ability to deploy social competences that engage with those situations. Whether an explanation in terms of behavioral rules (Ruffman and Perner 2005) can be vindicated or not, or whether we should think of such social competences in terms of a behavioral abstraction model or enactive perceptual abilities, it seems likely that the more parsimonious explanation will have to build on the relevant resources developed in processes that are closer to the perceptual and interaction processes of primary and secondary intersubjectivity than to metarepresentational and mentalizing abilities, the purported presence of which in young infants continues to surprise even theory theorists.

2 Chimpanzees: The Behavioral Abstraction Hypothesis

Chimpanzees (like humans and other animals) manifest a basic form of practical understanding of the behavior of others. We have argued that chimpanzees, confronted with the particular behavior of another chimp (e.g., pursing lips, bristling hair, etc.) are able to represent this behavior in terms of a more abstract interpretation (e.g., threat; Povinelli and Vonk 2003). Seeing a 'threat display', the chimpanzee likely has a sense of the kinds of behavior that will follow ('charging', 'being hit', etc.). We have referred to this as the 'behavioral abstraction hypothesis'. It holds that chimpanzees: (a) construct abstract categories of behavior, (b) predict future behaviors following from past behaviors, and (c) adjust their own behavior accordingly. Note that this idea differs sharply from the "behavioral rules" conception of Ruffman and Perner (2005) discussed by Shaun Gallagher in the previous section. We postulate that chimpanzees represent abstract categories of behavior and the networks of possible subsequent behaviors and reactions to their own behaviors. These are not rules, so much as possibilities for action weighed against the likelihood of certain reactions, and in this way at least, bears a strong resemblance to the enactive

perception model outline by Shaun Gallagher above. There may be important differences, as well, which we explore in the final section of this essay.¹

In contrast to the behavioral abstraction view, ToM-like theories suggest that chimpanzees construe behavior in terms of mental states. This requires the chimpanzee to infer beyond the behavior to the other's mental state and, accordingly, cope with both behaviors and mental states. Faced with the gaze of another, for example, the chimpanzee would have to represent this as both a behavioral code (involving abstracted spatio-temporal invariances) and a mental significance (e.g., that the other has a visual mental experience and is 'seeing', something). This ability to infer a mental state (i.e., a theory of mind, ToM) is a form of higher-order relational reasoning manifestly obvious in human behavior and cognition (see Penn et al. 2008). In this case, this process can be construed as involving the detection of something that is not directly perceivable in the behavior of the other – in this case, what we call a hypothetical mental state. In our folk psychology, we frequently recruit such mental states to explain overt visible behavior. Less frequently, we may use our recognition of the relations among relations (our representation of something like a mental state) to predict future behavior (see Penn et al. 2008).

In many cases, however, reasoning about mental states is redundant with the prior detection of abstract behavioral categories, since all of the information needed to predict future behavior is already contained in the perceivable behavior. In circumstances where a subordinate chimp will reach for food when they have seen the dominant chimp witness the hiding, but will not reach for food when they did not see the dominant chimp witness the hiding (see Hare et al. 2000, 2001), one can argue that the behavioral understanding may be sufficient: 'Don't go after the food if the dominant was oriented towards it'. One doesn't need to add the additional ToM clause: 'because he has seen it, and therefore knows where it is.' (Although we conveniently write this using human language, this phrase simply summarizes the output of a dynamic system that assesses likely outcomes given the external and internal inputs to the system, and selects the one most likely to achieve the motivation currently of highest value. Contrary to the opinion of some, this is in no way an "inflexible rule".)

In a further 'control' condition both the subordinate and dominant saw the food. Subsequently the location of the food was switched. When the subordinate but not the dominant saw the switch the subordinates were more likely to reach for the food than when the dominant also saw the switch. Again, it could be argued, that the behavioral rule is sufficient: 'He was present and facing the food when it was placed where it is now, so he is likely to go after it'. Do we need an extra step – 'he *believes* the food is

¹ One very important difference between these models concerns the language used to describe the underlying systems that implement these abilities. Our behavioral abstraction model is subsumed by a more global relational reinterpretation hypothesis which argues that humans uniquely implement a system (or systems) for higher-order role-based relational reasoning that accounts for uniquely human capacities in many domains of cognition (see below; Penn et al. 2008). At the same time it posits that in addition to sensory inputs, feedback from perception–action loops, autonomic–visceral states, the physical structure and capabilities of the organism's body and all the other many variables that influence the actions of situated, embodied, biological agents in the wild, much of the basic understanding of behavior characteristic of infants and chimpanzees (and adults most of the time!) as animals is supported by first-order perceptual-based relational reasoning. The enactive perception view would view these latter descriptions as excessively "cognitive" and would look for an explanation in terms of far more "embodied" understanding. These issues are discussed in more detail in the final section of this paper.

there’ – to help us predict his behavior? Or, ‘He was not present when the food was placed where it is now – so he didn’t see, therefore he doesn’t know – so he is less likely to go after it.’ On the one hand, according to the behavioral account, these references to hidden mental states do not operate like hidden premises necessary to arrive at the conclusion; they rather seem like redundant premises or unneeded extra steps. On the other hand, the behavioral account does require a hidden premise; namely, the behavioral rule, the result of some process of abstraction. In this respect, the behavioral rule hypothesis, as much as the ToM account, understands the process to involve inference. The ToM hypothesis simply adds a step (indicated in *italics*) to the inference:

Behavioral Rule Route	ToM Route
Behavior observation	Behavior observation
↓	↓
Behavioral abstraction	Behavioral abstraction
↓	↓
Inference to prediction	<Inference to mental state>
	↓
	Inference to prediction

Povinelli and Vonk (2003) argue that typical experiments are set up in a way that gives the chimpanzee access to behavioral rules and will not be able to establish whether the extra inferential step, i.e., ToM, is present (see Penn and Povinelli 2007). Thus, Povinelli and Vonk (2003) favor the more parsimonious explanation that does not require ToM:

Although [behavior or behavioral abstraction] may do much, if not most, of the actual work in supporting our behavioral interactions with others (and hence ought to be a greater focal point for research), if [some kind of mindreading] is not present, then we have no business invoking the phrase ‘theory of mind’ (see Premack and Woodruff 1978). (Povinelli and Vonk 2003, p. 157)

3 Continuity or Qualitative Difference?

We have concentrated on specific social-cognitive capabilities that appear to be somewhat and in some degree continuous among chimpanzees, human infants, and adult humans. Yet, in regard to social cognition more generally we know that there are clear and significant differences between chimpanzees and humans. Several questions come to the fore: How much continuity is there? How much is qualitatively different? And where precisely does the difference lie?

There is general agreement that there is something uniquely human about our ability to represent and reason about our own and others’ mental states (e.g., Tomasello et al. 2005). Most linguists and psycholinguists argue that there is a significant discontinuity between human and nonhuman forms of communication (e.g., Chomsky 1980; Jackendoff 2002; Pinker 1994). Tomasello and Rakoczy (2003, p. 121) argue that the ability to participate in cultural activities with shared goals and intentions is uniquely human, even if the human infant’s other cognitive skills do not

differ very much from the ape (Tomasello et al. 2003, 2005). Even if we say that there is, to a point, a continuity between human and nonhuman animals' abilities to learn about and act on the perceptual relations between events, objects, and encountered conspecifics, it is also right to say that humans are capable of reinterpreting certain higher-order relations between these perceptual relations in a systematic and inferentially productive fashion (see Penn et al. 2008). Only humans can go further in reflective metacognitive practices in which they are able to form general categories, find analogies, draw inferences, discover abstract functional regularities, reflectively consider their own cognitive processes and postulate the existence of, and reason about, mental states in others.

Clearly then, there are times when humans need and are able to move beyond what is perceptually available in the bodily movements, gestures, and facial expressions of others and to infer, simulate, and reason about mental states. This is not all the time, however, or even most of the time in regard to our everyday relations with others. In circumstances that are more ambiguous and less certain, or in situations where the other person presents puzzling behavior, we may indeed require theoretical inference or simulation skills, or more basically, the narrative competence from which such folk psychological skills develop (Gallagher and Hutto 2008; Hutto 2008). At a certain age, likely above 3 years, humans are able to reason about the motivations of other agents and their subsequent behavior (see Penn and Povinelli 2007 for a more extensive discussion of this point). Moreover, humans are also capable of simulating the mental states of others, as well as their actions and potential joint actions and interactions with others. Humans engage in these practices at least some of the time. But we see no reason to think that chimpanzees or young infants are capable of this, or need to engage in such processes to gain a good pragmatic and working understanding of the others with whom they interact and who inhabit the same world.

How we move from our social cognitive abilities that are grounded on enactive perceptual and interactional practices to higher-order abilities that involve the representation of mental states is still an open question. Whether it can be enactive all the way up (as well as all the way down), that is, an account worked out fully in terms of embodied and externalist principles (see e.g., Barrett 2008; Thompson 2007), to what extent it depends on narrative competence (Hutto 2008), or whether one needs to appeal to computational-representational models (like the relational reinterpretation hypothesis proposed by Penn et al. 2008) – we leave this as an issue to be addressed in ongoing debates.

In what follows, we attempt a discussion to see if it is possible to articulate the critical points on which some of these future debates might pivot. In doing so, we attempt to understand the similarities and differences between our two approaches for understanding the early social competences of infants, as well as some analogous competences in chimpanzees.

4 Enactive Perception and the Behavioral Abstraction Accounts: A Debate

Daniel Povinelli: Shaun, let me start with a question for you. Do you see the enactive account as an alternative to our behavioral abstraction hypothesis?

Doesn't the enactive account still need to grant apes/infants the ability to segment the optic flow into meaningful chunks? That's all we need to get the behavioral abstraction idea going, I think. No matter what language we use, the infant/ape still needs to have a way of keeping track of (1) meaningful units of the behavior of others and (2) meaningful units of the behavior of self, and (3) articulating them in "intelligent" (i.e., flexible) ways depending on context and background information and their current motivations. There would seem to be overwhelming evidence that both apes and infants are fully capable of first-order, goal-directed relational reasoning. This means that they must be able to flexibly substitute different actions depending on the goal, and use the same actions to achieve different goals. Our hypothesis amounts to a theory of how the perception of action leads to the selection of appropriate actions. But a selection process is still necessary on the enactive view, correct? You would agree that the existence of the perception action loop in no way eliminates this selection process and therefore the need for decisions among actions? Although we frequently call these decisions "inferences", I am not sure if this is a substantive or terminological difference between the enactive perception model and our own.

Shaun Gallagher: I agree that apes/infants can perceptually differentiate intentional, goal-directed actions – infants around 6 months of age start to perceive grasping as goal directed, and at 10–11 months they are able to parse some kinds of continuous action according to intentional boundaries (Baldwin and Baird 2001; Baird and Baldwin 2001; Woodward, and Sommerville 2000). And they start to perceive various movements of the head, the mouth, the hands, and more general body movements as meaningful, goal-directed movements (Senju et al. 2006). I would characterize this as a perceptual ability; you may want to say this involves behavioral abstraction. What the enactive position adds to the behavioral abstraction position concerns the nature of the meaning that I see in the other's actions. The other's actions have meaning for me in terms of how I may be able to interact with her – something that I don't have to calculate consciously, but rather something that registers at a bodily level. Mirror neurons may be involved, but as preparatory for action rather than as simulating or going into matching states. Yes, this still implies "the ability to segment the optic flow into meaningful chunks" (as you state). But the 'meaningful' aspect relates to one's own potential to interact with the other – which is something defined precisely in terms of one's own bodily state and spatial location. That is, the meaning isn't an abstract or theoretical or intellectually rational thing (as if we just solved a puzzle), but a matter of practical or pragmatic (and specifically social) reason. As you say, "the infant/ape still needs to have a way of keeping track of (1) meaningful units of the behavior of others and (2) meaningful units of the behavior of self, and articulating them in "intelligent" (i.e., flexible) ways depending on context and background information." This is exactly right I think. The only issue is what defines meaning – and the enactive view is that the meaning is defined by the affordances offered by the other. I think this is consistent with your view, but offers a specification about the meaning. As you say, your theory is, "essentially, a theory of how perception of action leads to

the selection of appropriate actions.” The enactive idea defines what ‘appropriate’ means – i.e., appropriate in relation to what the action affords and what my capabilities are for responding – even if I’m only observing and my actions are inhibited. I would avoid the term ‘inference’ simply because it sounds too “in the head,” and the enactive view suggests that a lot of the work is already accomplished in the dynamics of body-environment (not inferences in the head) – or in the case of social engagement, in the interaction itself (De Jaegher et al. 2010) or how my body gears up for potential interaction.

DJP: The behavioral abstraction model doesn’t suggest the selection process is a reflective one. The language “this is possible if x ” or “this is unlikely if y ” therefore “do z ” is just a shorthand for describing the information that must be present in the brain/body of the agent and some kind of weighted action selection process. And aren’t I correct in thinking that the “meaning” the enactive view seeks to specify “in terms of the agent’s current bodily states, spatial position etc.” still needs to be implemented in terms of some kind of “selection” process? We might disagree about how much of this selection is “embodied” in terms of the peripheral nervous system (PNS) versus more “centrally” located in terms of the central nervous system (CNS), but in principle I don’t see why one is more “theoretical” than the other. It’s still a decision process – a selection process involving choosing behaviors (actions) that are possible given what I register about the world. The “consciousness” of these selections seems orthogonal to the issue.

SG: I think you are right about the first part of this, as long as we don’t end up saying “... and then a selection process happens in a set of inferences that take place in the brain”. Not to deny brain involvement; it is involved as an important part of the system, which includes PNS and which receives important specification from the physical and social context of action and interaction. In some (maybe most) contexts what I see in the embodied movement of the other may lead me to (or draw me into) the right response – so the selection process (which is rational but more pragmatic than theoretical) is distributed to some extent. In other contexts it may be a matter of reflective reason (and more in the head).

DJP: Okay, but let me clarify that I don’t care what we call the selection process. I think it’s clear that my colleagues and I don’t think that any kind of metarepresentation is involved in the chimp’s ability to make “inferences” using behavioral abstractions. Turning back to the enactive view, I can understand (in general) how you might construe results like Southgate and colleagues’ from the standpoint of the enactive perception or behavioral abstraction models. The specifics may not be completely laid out yet, but (in general), in this case, the actions and situations are related to (1) pointing, (2) naming, and (3) retrieving. Setting aside how exactly all of this would be described in embodied and/or representational (gulp!) terms, these particular behaviors surely have both an ontogenetic and evolutionary history in which the perception/representation of affordances for actions and interactions of actions could be “stored” somehow

in the body/brain of the observer. But what would you say about the recent study by Senju et al. (2011)? In this study, infants first were given first-person experience wearing one of two kinds of blindfolds. One group of infants experienced “trick” blindfolds that could be seen through even though others could not see their eyes (at least I assume this, if not the study is not valid). The other group experienced “standard” blindfolds, which could not be seen through. While wearing each of these they were asked to name various pictures and toys. After this, they observed a video involving other agents watching toys hidden in location A. Then the agents put on a blindfold that ostensibly matched the blindfold type the infant had experienced, and the objects were switched to location B. The infants who had experienced the standard blindfolds, saw a puppet switch the toy to location B. The infants who had experienced the trick blindfolds saw the puppet remove the toy from the scene. The standard group glanced to location A first, whereas the trick blindfold group glanced at random. The experimenters interpret this to mean that infants who had experienced the trick blindfold understood that these agents could <see> the switch. I’ll use this notation (<see>) to denote when I’m talking about the folk notion of a general experiential aspect of seeing that links many contexts and particular kinds of experiences of “seeing.” Now, to be clear, I don’t defend the exact design of Senju et al. (see Povinelli and Vonk 2004; Penn and Povinelli 2007; Vonk and Povinelli 2011, for our design). In my estimation, the toy should have been switched to location B for the trick group. Under that design, the model that posits that they represent <seeing> would then predict that these infants would look to B, whereas the other group to A. We don’t know what the infants would do, of course, but to advance the conceptual issues, do you think that such a task (hereafter “a Senju-style” task) has this resolving power?

SG: Senju et al. seem to think that understanding another person's action based on behavioral cues is something that happens in thin air and is totally independent of the subject's experience. (I assume that is not the way you think of it.) The enactive version of this is that yes, of course infants rely on their previous experience with the blindfold. That experience attunes their pragmatic grasp of the sensory-motor contingencies afforded by the blindfold. They accordingly recognize (know, have a sense of) what the blindfold affords or does not afford at the level of embodied perceptual comportment and when they see the other person wearing the blindfold they grasp that affordance – without having to attribute a mental state standing behind that comportment. The infant, familiar with the non-opaque blindfold, knowing that it does not blind the person who wears it, can see that the person sees what is going on. I don't think an inference to a mental state beyond that is necessary. Where one draws the line between *see* in some behavioral sense, and <see> as experiential state– which I take it, for you means a mental state – is a good question. I don’t think the infant makes a distinction. I don’t think human adults normally make a distinction. When you ask someone whether they see the jazz quartet on stage, they don’t introspectively consult their mental states – they look on stage. Why should it be so different if we ask someone if the person next to her sees the quartet. They look at the stage and they look at the other person and they see the line of sight – they

don't have to think of it in terms of some kind of internal state in the other person's head.

DJP: Of course I agree with your latter point. We have argued for many years that human adults rarely make this distinction (Povinelli 2000, 2012, for example). But I do not find this analogous to the Senju-style approach. When I report that my buddy can see his friend "seeing" the jazz quartet, his answer is a quick linguistic response based on registering the geometry between his friend's eyes and the stage. The question is, *Can* (not *must*) my buddy also represent an additional fact about the other person: that they are having a subjective visual experience that may have additional, causal implications for their behavior? The answer, I take it, is yes. Let me back up for a moment, just to be clear, and ask what does "embodied perceptual comportment" mean here, precisely?

SG: There may be some redundancy here, if one thinks that all comportment (behavior) is embodied. The words 'embodied' and 'comportment' are intended to emphasize that perception is not a pure mental state but is something accomplished bodily, involving sense organs, posture, and position, as well as brain processes. So, perception is not entirely a hidden internal state – which is not to deny an experiential component. The question is whether we need to replicate or infer that experiential component (mental state?) in order to grasp that someone is perceiving. I think we agree that we don't need to do so; at some point, however, humans are able to do so, if the situation calls for it.

DJP: I agree that perception resides in a highly distributed fashion throughout the body. In fact, we show this quite nicely in our latest large-scale empirical project involving the levels of the chimpanzee's understanding of the physical dimension of "weight" (Povinelli 2012). But as you know, we've long-advocated a Senju-style task as a valid, minimal assay of the ability to represent mental states (in this case, the experiential aspect of perceptual states: <seeing>, <hearing>, etc.), in contrast to the state of the system itself. The chimpanzees we tested could not successfully cope with this task (Vonk and Povinelli 2011). Of course, I agree that the infant's abilities to reason about the actions of others are dependent on their own first-person experience. But does your view imply that the infants "know/have a sense" of an internal perceptual state of the other? A state distinguishable in some way from head orientation, eye direction, etc.?

SG: I don't know if infants have a sense of the 'internal' perceptual state of the other, but I think the evidence suggests they have a sense of the [external] embodied perceptual state of the other, and this is sufficient for understanding.

DJP: Perhaps I shouldn't have said 'internal.' What I mean is, do you find evidence that they represent more than the individual perceptual relations themselves, or as Derek Penn and Keith Holyoak and I have outlined (see main text), do they recognize/compute/identify the commonalities among perceptually distinct relations – in this case, the relation between my first-person

experience of having worn standard versus trick blindfolds and your reactions to many other perceptually disparate situations when you are wearing them?

SG: That's one way of putting it, and I don't want to insist that you have to use enactive vocabulary to say it. I think they likely do see commonalities (infants can imitate, after all) and relations, and more than that. Do they start to abstract and categorize these relations? At some point I think they are capable of that, but I think we would have to sort out what categorization means on a perceptual level versus a higher-order level that may depend on language ability. I don't think that there is, for example, a visual perceptual state distinguishable from head orientation, eye direction, etc. This does not mean that these are mere behaviors; rather, because they happen in specific contexts it means that these are meaningful behaviors that involve experiences. That doesn't mean that we have to infer anything about those experiences to make sense out of what we see in the other's behavior. I think you agree with that.

DJP: When you say that the 1st person experience "attunes" them to the SM contingencies afforded by the blindfolds, are these observable in the other person?

SG: Let's leave the blindfold out of it for a minute. We could say that the infants' first-person experiences starts to attune them to the SM contingencies afforded by the other person's position, the way their head is turned, whether their eyes are open or closed. Quite simply, if the other person is not looking in a certain direction, she will not see the object that lies in that direction. The blindfold is just an add-on to that. I can observe the person's position, the way her head is turned, or whether her eyes are open or closed – and I can observe that they are wearing a blindfold. These aspects are observable/perceivable.

DJP: Yes. No dispute that these latter things are in the province of perception.

SG: And on the basis of my previous experience with the blindfold, I can grasp whether they see the situation or not. Should we call this an inference? I want to resist that, but even if we do call it an inference, it's not necessarily an inference to a mental state that is hidden away in the other person's head. One's perception or lack of it is a function of eyes, blindfolds, physical positions – it likely has an experiential (mental) aspect to it (there is something it is like to perceive something), but I don't have to worry about that to understand whether you see something or not.

DJP: But with respect to the Senju-style experiments that we have advocated, I don't see how it follows that based on my previous experience with the blindfolds I can "grasp" whether they can <see> or not – if by "see" you mean a generalized mental state. And I think the folk psychological construct of <seeing> is what is at stake in a Senju-style study. If the infants are making predictions about behavioral and physical disposition alone, how do they bridge the relational gulf between what the infant experienced (wearing the two types

of blindfolds and naming objects) to the predictions they are making by observing in the video (involving what another agent will do when the locations of toys are switched)? If the infants only know what sorts of behaviors are afforded by the blindfold type, then why (in the opaque blindfolds condition) do infants exhibit a structured bias in their gazing (i.e., in the place where the toy had been)? Note that in the hypothetical outcome of the properly designed study, this manipulation does not cause them to look randomly. Rather the infants in this condition would be making an inference that goes beyond “knowing that the actor did not see [read: head/eyes oriented toward] the switch” – it implies that they recognize that the manipulations structure the actor’s next action in a particular manner. In particular, that the “lack of *<seeing>* the switch” will lead to a previous, default action. To me, this implies an understanding of a context-independent aspect of seeing – a thing distinguishable from any particular overt indicators of “seeing”.

SG: Why? I’m not sure I get that. I know that one can put this into a folk psychological formula: mental states motivate action. So, to predict what the action will be (where they will look), I have to know the other person’s mental state (their perception). But again, their perception is something that is right there in front of me – in their embodied comportment – their location vis-a-vis the object; their open or shut eyes, a blindfold that allows vision or doesn’t. Does one have to worry about or make inferences about the other person’s mental states to work this out?

DJP: But “open vs. shut eyes” seems very different from “agent experiences *<seeing>* vs. agent does not experience *<seeing>*”. The former is within the province of perception, the latter is not. I cannot perceive your experience of *<seeing>*. I can perceive you wearing the blindfold types, but not your experience while wearing them. Let’s try this. Let A = an agent perceiving (via their sensory systems) the dispositions of the sensory systems of others, AND the agent understanding all related potential behaviors that are likely or not likely to follow or be possible from these dispositions, and let B = an agent representing/simulating/keeping track of the general subjective consequences of the perceivable dispositions of the sensory systems of others (a subjective qualitative state they might experience such as a feeling or a general visual experience). Without debating about the frequency or daily importance of either A or B, do you agree that both are a part of human cognition?

SG: Yes, I agree that both are possible.

DJP: Can you say what you take as a criterial test of whether B has occurred?

SG: I’m not an experimentalist, so it’s difficult for me to say. I agree with what you say in Penn and Povinelli (2007), that it is difficult to test for B when there is normally a lot of A around. If we could beg the question a little, then in a specific case of high-functioning autism where (at least on one reading) the capacities of primary and secondary intersubjectivity are absent, the autistic

person reports using logical inference and reasoning to figure out what the other person is thinking. That would be B without A. One would have to do a series of tests and depend somewhat on the reports of the autistic subject. Otherwise, I'm not sure how one could block the influence of A.

DJP: That example does not seem like B without A. It seems more like arriving at B by effortfully triangulating on the relations between A and B (or at least an analog of B). The autistic individual would still need to track the behavioral abstractions. More generally, we have argued that in the real world, A is necessary for B but not vice versa.

SG: Sorry, I took A to involve the perception of something richer than pure mechanical behavior – and maybe this is a telling difference in our understanding of the Senju study – although more generally I think we agree. I took A to mean something like the perception of a rich and meaningful and contextualized behavior (call this A+) – the sort of thing one comes to understand in the interactive and perceptual abilities of primary and secondary intersubjectivity. This, or some important part or degree of it, is what I take to be missing in autism. The autistic person doesn't have A+, he has A-, an impoverished perception, or a perception of a close to mechanical kind of behavior, and on that basis, if he is on the higher-functioning end of the spectrum, he tries to puzzle things out via B in order to make sense of A-, since for him there is no sense implicit in A-. In other words, he has to mindread. So you're right, in this example it's not B without A, but it's B with A-. And what we want to say, I think, is that if we have A+, we don't have to go to B, where B consists of trying to grasp the other's mental states or 'internal' experiences of any sort. So in the case of the blindfold, I don't have to bridge a gap or a gulf between what's in the province of perception and what it not. I agree that I can't literally see your visual experience. I'm tempted to say that I can see *that* you visually experience something (or don't) by seeing your open (or shut) eyes, your wearing of the blindfold plus my prior experience with the blindfold – but this doesn't mean I have to perceive your experience.

DJP: I have a different take on the autism issue, but I do not think it is relevant here. In general, I mean "A" in the way that you describe, although I am not completely certain what you mean by "pure mechanical". I would say that I think chimpanzees and young infants can act upon categories of actions and select appropriate responses because they somehow "represent" the probabilistic linkages among them. I would simply seek to specify (break down) the content of the perception of/representation of the "rich and meaningful and contextualized behavior" to which you refer. To be sure, in the case of autism *some* (but not all) of the linkages may not be fully understood so we can label that A-. In any event, my question concerning what you would find criterial of a situation in which B is present in addition to A, was because of your earlier statements that "perception is not *entirely* [my emphasis] a hidden internal state – which is not to deny an experiential component. The question is whether we need to replicate or infer

that experiential component (mental state?) in order to grasp that someone is perceiving.” My own inclination is to say ‘no’ – if by “grasp that someone is perceiving” you mean keep track of the aspects of you that correspond to you <seeing> something. However, this in no way means that, in addition, I understanding someone perceiving as someone having an internal subjective experience. And that’s all that I see is at stake here.

SG: I think we agree on that.

DJP: But earlier you stated: “I don’t think that there is a perceptual state distinguishable from head orientation, eye direction, etc. This does not mean that these are mere behaviors; rather, it means that these are meaningful behaviors that involve experiences that we don’t have to infer to make sense out of what we see.” So doesn’t this imply you do not accept the existence of B?

SG: I accept that B (i.e., trying to grasp other people’s mental states, or explain their behavior in terms of such mental states) is possible but rare and, as you indicate in Penn and Povinelli (2007), would be redundant, since the work is already being done by A+ in most everyday cases. A+ is sufficient to get to the (often rich) meaning of the behavior.

DJP: Yes, I think we both agree that the use of B (which includes A) is rare in everyday social interactions. And that it follows that much of the time we and chimps are operating on the basis of something like A alone. Now, in terms of A getting to the “meaning of behavior” would you equate “meaning” with “network of relations”?

SG: You mean the set of relations it is confronted with in the experimental manipulations? Yes.

DJP: I do mean that, but I also mean the information of the goal-directed, first-order perceptual relations the infant carries with them in their bodies and brains from both their basic starting biology and learning. In any event, I am positing that a properly designed Senju-style study is not a case of A (or A+) alone.

SG: I’m trying to understand why it is not a case of A+ alone, or why you would have me start thinking of the experiential consequences of wearing one or the other blindfold.

DJP: Let’s say infants “perceive someone perceiving” as in A+. We both already agree they do this. I have also argued that simply slapping a blindfold on someone and showing that infants respond differently is emphatically not evidence of B because what the infant has come to understand as “visually perceiving” is not met. Namely, the eyes are obscured/missing. Note that in the methodology at hand, the question becomes not whether the agents wearing one set of blindfolds are perceived as “different” from agents wearing the other – they will on the A+ account.

SG: So far so good.

DJP: The question is, how can we explain a structured response of the infants in a Senju-style study? Namely, that if they saw someone wearing the trick blindfolds they predict they will look where the toy was relocated. This makes sense if, in addition to A, infants also are capable of B. But I cannot see how that emerges from A alone.

SG: As I understand Senju et al., the infant wears an opaque blindfold and learns that the blindfold prevents one from seeing (much as a physical barrier does). An infant knows it can't see through a wall, and now it learns that it can't see through an opaque blindfold. When that infant sees an agent with a blindfold on, she assumes that the agent can't see what happens, and so expects the agent to look in the original location. A different infant wears a transparent blindfold and learns that the blindfold does not prevent seeing – it still affords seeing; when that infant sees an agent with a blindfold on, she assumes the agent can still see and so does not expect the agent to look in either location since they perceived that the toy was removed. The infant is able to understand the situation in terms of what it affords or doesn't afford. This is an enactive view defended recently by de Bruin et al. (2011) – the idea that we can perceive the other's affordances within specific situations. If so, there doesn't seem to be a problem that would cause the infant to start concerning itself about B. This is about the infant's perception (informed by past experience – do Perner and Ruffman, or anyone claim that we rely *solely* on behavioral clues?). The infant perceives the other agent being in a position to perceive in a set of physical circumstances that either affords or doesn't afford seeing what's going on.

DJP: What does “perceive perceiving” mean here? I maintain that for chimpanzees, they perceive a set of behavioral categories that correspond to a set of behavioral dispositions. If they “perceive perceiving” in some other way, it would seem to be as a higher-order variable that bridges perceptual discontinuity between the chimp's experience and the other's experience. I do not perceive your subjective visual experience. We agree on that, correct?

SG: Right, I do not perceive your experience. I don't think we perceive categories either. I perceive the situation we're in, I see your gaze, the position of your head, your line of sight, your actions and so forth. I also have some awareness of what we have been doing and where we're heading, especially if we have been interacting. In such cases (which may be most cases) I do not add up these various perceptual items as if they were premises in an argument and then conclude that “you are having a visual experience.” Rather my perceptions are feeding my responses or potential responses to you, and I perceive your actions in terms of our continuing interactions. In this there is all the meaning that I need for understanding you, and I don't have to go to B. Many of the experiments arrange things so that I'm a third-person observer, and in such cases I may have more or less information to work with. In this situation, and depending on the complexity of your action, I may have to go beyond what I

see of your behavior and try to sort out some story by way of B. But for purposes of understanding your behavior, or what this means, in most circumstances I don't have to go that far. Senju et al. jump to the idea that the infant infers a false belief in the same way that Baillargeon et al. jump to that idea. It's the standard way of interpreting this result. But I don't see why the infant needs to infer a belief at all. I don't see why B is necessarily involved since there is plenty of information available via A+

DJP: I don't agree on the structural similarity between the two, Baillargeon and a Senju-style test. In the Baillargeon case, there is ample opportunity to learn that people tend to grasp where they are looking. And they have had plenty of extensive prior history of observing the contingencies among the geometrical layout of eyes and opaque barriers and resulting actions.

SG: How is an opaque blindfold any different from a barrier that blocks my view? If I see that an agent is in such a position so as not to be able to see X because there is a barrier in the way, is that different from seeing that an agent has a blindfold barrier on his eyes? How do I know that the agent can't see around or through a barrier?

DJP: The issue isn't blindfold vs. barrier. This is not what distinguishes the design of a Senju-style study. It could have been two barrier types. The issue is how the infant tracks the relations between *types* of blindfold (yielding different first-person experiences: "darkness" vs. "visual experience") that they have no prior history with and a very limited set of behavioral propensities. There is no question that the infant experiences that its own behavior is manifestly altered by blindfold type. But this is not the same set of relations it is confronted with in the experimental manipulations where others are observed wearing the two blindfolds. So the real issue involves the causal work either being done or not being done by the infant's understanding of the experiential aspect of <seeing>. In Penn and Povinelli (2007), we argue examine all of the possible variables we can think of and show that short of positing B, this problem cannot be solved (see pp. 737–739). I take it you disagree with our formulation of this problem? If so, how?

SG: In that paper you state: "In this context, let us examine more closely the data available to a subject lacking an $f(\text{ToM})$. Such a subject would be limited to r-states about his own manifest behaviour while wearing the opaque visor (e.g. 'I stumbled around while wearing the red visor') and occurrent p-states about the experimenter (e.g. 'she is wearing a red visor'). However, a subject lacking an $f(\text{ToM})$ would not have access to r-states about his own internal cognitive states while wearing the visors (e.g. 'I was unable to see while wearing the red visor')." I take it this means that unless I have ToM capability, I do not have access to my own experience or r-states. I know that Carruthers and some others would say this, and they seem to mean that we can only infer what our own mental states are. I would reject this on phenomenological grounds. But even if it's right with respect to one's own belief states (that would be the most

moderate version) is it really the case that when I perceive something my access to the experiential aspect of that perception is only by inference or some ToM procedure? I put on the red visor, and I can't see. Do I then have to infer "I was unable to see while wearing the red visor"?

DJP: I agree you are "aware of the experience" but this is not the same as saying that you can identify *<seeing>* – if by *<seeing>* we mean something other than a given experience. For example, we do not take this to be different from having a physical experience such as the effort to lift an object that can be related of some folk physical property (like *<weight>*), and even tracking that experience in future lifts, and yet not ascribing some higher-order variable like weight to objects that explains a wide range of effects weight might have in the world. We take the case of *<seeing>* to be the same, roughly speaking: I can have a visual experience without identifying it elsewhere as a common cause of many disparate phenomena (see Penn et al. 2008).

SG: I'm not sure I follow you on this. Are you still defining *<seeing>* as the experiential aspect? If so, what does it mean to say that 'by *<seeing>* we mean something other than a given experience'? Isn't *<seeing>* something that I just experience, and am I not self-aware when I am living through that experience? Someone who accepts the ToM idea about first-person experience would think there is a gap to be filled. I don't think that you do accept the ToM idea about first-person experience. The phenomenology here is not that (1) I experience x, and then (2) I represent that experience. The phenomenology is rather that in experiencing x I am pre-reflectively (or ecologically, if you think of Gibson) self-aware of that experience. The experience is co-presented with the x. No representation is needed for me to be aware of my own experience. So I would say, I have an experience of not seeing while wearing the opaque blindfold (or of not being able to see through a barrier), and to have learned from that, I don't need ToM. My vision is blocked, and I have an immediate sense that this limits my behavior and the affordances the world offers me.

DJP: I would say you have the experience, period. But what I mean by picking out *<seeing>* is that you can understand how a wide range of disparate perceptual experiences (in both the self and other) are all part of something that, as you say, is not perceivable. This is what we take to be evidence of higher-order, role-based, relational reasoning (Penn et al. 2008). Theory of mind is just one example of this (we argue) uniquely human ability. Thus, you may or may not be able (somehow) to understand that a given experience connects to a broad range of other situations such that there is something you want to label using folk psychology as *<seeing>*. I would say that the "experiential" aspect of seeing is what our folk psychology picks up on and labels as being common to many very different instances. Now, I note that you use the term "learn" here. I interpret this to mean that I could learn two relations. Relation 1 = wearing standard blindfolds leads to "experiencing darkness". Separately, Relation 2 = wearing trick blindfolds leads to "experiencing non-darkness".

SG: Part of what it means to learn involves being able to generalize. I then see that the opaque blindfold (or some barrier) is blocking the view of some other agent who is wearing it. Don't I then have a sense that this limits that agent's behavior and their affordances, including some of the affordances that pertain to interacting with me?

DJP: Yes. But note the ambiguity is the phrase "blocking the view". You mean: "someone with standard blindfolds between their eyes and objects in the world", correct? And this does not require me to track my visual experience as a general category of subjective experience that can be born out in many perceptual distinct relations.

SG: Yes, but do I need to worry about their mental states in order to get this far?

DJP: I don't think so. But in a properly designed Senju-style experiment, I would think that the question is not, Is it possible for infants to know that others will act in a different way depending on the infant's familiarity with blindfold type 1 or 2? We both agree that their own first-person experience allows them to identify (predict) that their experiences are different wearing the two blindfolds. The question is: can they immediately know that another agent wearing standard blindfolds will cause them to look in location A? This is a relation they have not learned from their own experience with the various blindfolds. And I do not see how it is given in perception. The only thing that is given in perception is what they perceived while they were wearing them (e.g., standard: "darkness", "stumbling around", "inability to answer questions about objects", etc. trick: "experience", "not stumbling", "ability to name objects") and what they perceived while you were wearing them (trick and standard: "covering your eyes", etc.). They did not perceive themselves "not <seeing>" nor did they perceive the other "not <seeing>". Or if your definition of perception includes the perceiving of subjective experience of <seeing> as something very general that connects across many instances (your experiential aspect of seeing) then we have just moved the information problem under the heading of perception. As I understand it, this requires the infant to track something other than many, separate, goal-directed primary perceptual relations. They would be tracking the relation between and among these relations. And this, we argue, constitutes a minimal ToM (Penn and Povinelli 2007).

Let me see if I can summarize the differences in our views. On your view, the infants in a Senju-style study perceive that standard blindfolds = "darkness", trick blindfolds = "subjective visual experience"

SG: I would say, in this shorthand, that trick blindfolds = being able to see the world. What I experience in this case is the world, not my subjective visual experience as such.

DJP: You mean, trick blindfolds = "the world as such"? Does "see" come into it because "being able to see the world" is a shorthand for us theoreticians to keep

track of the fact that the agent's visual experience presents itself in a manner different from the way hearing presents itself, for example?

SG: When the infant wears the trick blindfold she learns that it does not block her view; she is able to see the world. That's all I mean.

DJP: Okay. I would say that without a representation of <seeing> the infant learns "experience q." Now on your view, the infants also learn a set of contingencies between darkness and various things they can and can't do. With standard blindfolds I can "hear, reach out, wave my arms, awkwardly bump my hands into things", but I cannot "respond correctly when asked to name of toys, locate objects in space, etc.". *Mutatis mutandis* for the trick blindfolds. Now they confront others wearing these types of blindfolds. If I understand you, you believe that the infants can generalize from their behaviors and experience with these novel blindfolds, to the affordances of how the others will act in an unrelated (unrelated to what they learned about their own behavioral affordances while wearing the blindfold types) object switch paradigm, and they do this without explicitly taking into account the other's subjective experience of <seeing>.

SG: Not entirely unrelated. I learn from wearing the standard blindfold that I can't see when it covers my eyes. I learn something about blindfolds in this case. Blindfolds are like walls that I can't see through. Now I see someone else wearing a blindfold. I know how blindfolds work – what affordances they provide or don't provide – like I know how walls work. This aspect is related to every context in which I see someone wearing this kind of blindfold, including the context in which that person doesn't see a switch being made. Why I think the person will look in location A or not is related to the fact that I know they have seen or not seen (in the case of the blindfold) the switch. Knowing that they have not seen it doesn't necessarily mean I attribute to them the mental experience of <not seeing> the toy, as Senju et al. suggest; rather it means that I see they are wearing the blindfold, which I know blocks their view. *Mutatis mutandis* for the trick blindfolds and <seeing>.

DJP: I think that I am having a hard time grasping how infants can track the many instances of "seeing" (standard blindfolds = total darkness [experience *q*], opaque barrier = experiencing some but not all of the world [experience *r*], vantage point = experiencing an object from a point of view [experience *s*], transparent barrier = experiencing what is on the other side of a barrier [experience *t*]) as the same thing without having the ability to track the relation(s) among all these relations (Penn et al. 2008) – in other words, without representing something like <seeing>. On our theory, to do so would amount to a representation that is not perceptual and one that our folk psychology describes as <seeing>. In this particular case, for example, I am having a hard time following how the infant could generalize to infer/know/expect/understand the relevant affordances that are present in the object switch task. From their past learning, they know many affordances that flow from someone being in a

geometric position vis-à-vis a wall or having their eyes obscured etc. But it follows from what I said above that the same first-person experiences are not embedded in these relations, nor do they neatly match up with the same set of affordances. Now they enter a properly designed Senju-style study and learn a specific set of affordances for me that stem from a unique pair of blindfolds they have never experienced before. When you say they know that this or that “blocks their view”, what could that mean other than (1) it limits/allows a certain subset of affordances, or (2) they cannot <see>? If it’s the former, the affordances would not seem to transfer. I can clearly understand how the infants might think both groups of agents might select location A (based on their prior experience of seeing how people behave when their eyes are obscured, or when a barrier with certain definable properties are in a certain position relative to their eyes).

SG: I’m not sure it’s ‘either-or’. One of the affordances that the standard blindfold limits is <seeing> but, again, I don’t think that infants distinguish ‘seeing’ from <seeing>, either in their own case or in the case of someone else. The opaque blindfold means that whoever is wearing it has no visual access to the world.

DJP: Okay, I think I might need to reevaluate my views on this particular task in light of this discussion. In any event, let’s return to where we started. Although we default to different language to describe our conceptual analysis of the Senju-style task, in the end our temporary disagreement does not seem to stem from any inherent differences between the enactive model and the behavioral abstraction model. Let’s stipulate for a moment that your analysis is correct. If true, wouldn’t this simply mean that you have identified some particular abstract features of the disposition/behavior of others that the infants are (somehow) keeping track of in their bodies and brains? I use the phrase “abstract” here in the way we have been careful to use this term: perceptual abstractions/perceptual generalizations (see Penn et al. 2008; Povinelli 2012, Chapter 9). Do we agree on that?

SG: I agree that infants can keep track of abstract features, as you define the term. I don’t agree that these abstract features lead the infant to infer unperceived mental or experiential states or that they necessarily register for the infant as such. Or for the adult for that matter, although I don’t deny that at some point in development we gain ToM capability which we *can* use in some circumstances. But this, as I, and others, have argued elsewhere, depends not only on aspects of primary and secondary intersubjectivity, but also on the development of communicative and narrative competencies.

DJP: Thank you, Shaun. I remain agnostic on the empirical side of the question regarding infants. But I do agree there’s really good evidence that narrative competence is correlated with ToM competence. And finally, although I cannot yet say whether my body understands everything you are communicating to me, I am sure the answer is embedded somewhere inside me.

SG: Yes, I see what you mean.

References

- Baillargeon, R., R.M. Scott, and Z. He. 2010. False-belief understanding in infants. *Trends in Cognitive Sciences* 14(3): 110–118.
- Baird, J.A., and D.A. Baldwin. 2001. Making sense of human behavior: Action parsing and intentional inference. In *Intentions and intentionality: Foundations of social cognition*, ed. B.F. Malle, L.J. Moses, and D.A. Baldwin, 193–206. Cambridge: MIT.
- Baldwin, D.A., and J.A. Baird. 2001. Discerning intentions in dynamic human action. *Trends in Cognitive Science* 5(4): 171–178.
- Barrett, L. 2008. Out of their heads: Turning relational reinterpretation inside out. *The Behavioral and Brain Sciences* 31: 130–131.
- Buttelmann, D., M. Carpenter, and M. Tomasello. 2009. Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition* 112: 337–342.
- Carruthers, P. 2009. How we know our own minds: The relationship between mindreading and metacognition. *The Behavioral and Brain Sciences* 32(2): 121–182.
- Catmur, C., V. Walsh, and C. Heyes. 2007. Sensorimotor learning configures the human mirror system. *Current Biology* 17(17): 1527–1531.
- Chomsky, N. 1980. *Rules and representations*. New York: Columbia University Press.
- Csibra, G. 2005. Mirror neurons and action observation. Is simulation involved? ESF Interdisciplines. <http://www.interdisciplines.org/mirror/papers/>.
- de Bruin, L., D. Strijbos, and M. Slors. 2011. Early social cognition: Alternatives to implicit mindreading. *Review of Philosophy and Psychology* 2: 499–517.
- De Jaegher, H., E. Di Paolo, and S. Gallagher. 2010. Does social interaction constitute social cognition? *Trends in Cognitive Sciences* 14(10): 441–447.
- Dinstein, I., C. Thomas, M. Behrmann, and D.J. Heeger. 2008. A mirror up to nature. *Current Biology* 18(1): R13–R18.
- Dokic, J. 2002. Reply to Pierre Jacob. In *Simulation and knowledge of action*, ed. J. Dokic and J. Proust, 111–117. Amsterdam: John Benjamins.
- Gallagher, S. 2005. *How the body shapes the mind*. Oxford: Oxford University Press.
- Gallagher, S. 2008. Inference or interaction: Social cognition without precursors. *Philosophical Explorations* 11(3): 163–73.
- Gallagher, S., and D. Hutto. 2008. Understanding others through primary interaction and narrative practice. In *The shared mind: Perspectives on intersubjectivity*, ed. J. Zlatev, T. Racine, C. Sinha, and E. Itkonen, 17–38. Amsterdam: John Benjamins.
- Gallagher, S., and D. Zahavi. 2008. *The phenomenological mind*. London: Routledge.
- Goldman, A.I. 2006. *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. New York: Oxford University Press.
- Goldman, A., and C.S. Sripada. 2005. Simulationist models of face-based emotion recognition. *Cognition* 94: 193–213.
- Hare, B., J. Call, B. Agnetta, and M. Tomasello. 2000. Chimpanzees know what conspecifics do and do not see. *Animal Behaviour* 59: 771–785.
- Hare, B., J. Call, and M. Tomasello. 2001. Do chimpanzees know what conspecifics know? *Animal Behaviour* 61: 139–151.
- Herschbach, M. 2007. The phenomenological critics of folk psychology: The case of false belief. Paper presented at the Berkeley-Stanford-Davis Graduate Student Conference, 14 April 2007.
- Hurley, S.L. 1998. *Consciousness in action*. Cambridge: Harvard University Press.
- Hutto, D.D. 2008. *Folk psychological narratives: The socio-cultural basis of understanding reasons*. Cambridge: MIT.
- Jackendoff, R.S. 2002. *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.
- Merleau-Ponty, M. 1964. In *The primacy of perception*, ed. J.M. Edie. Evanston: Northwestern University Press.
- Newman-Norlund, R.D., M.L. Noordzij, R.G.J. Meulenbroek, and H. Bekkering. 2007. Exploring the brain basis of joint attention: Co-ordination of actions, goals and intentions. *Social Neuroscience* 2(1): 48–65.
- Noë, A. 2004. *Action in perception*. Cambridge: MIT.
- O'Regan, K., and A. Noë. 2001. A sensorimotor account of vision and visual consciousness. *The Behavioral and Brain Sciences* 23: 939–973.

- Onishi, K.H., and R. Baillargeon. 2005. Do 15-month-old infants understand false beliefs? *Science* 308: 255–258.
- Penn, D.C., and D.J. Povinelli. 2007. On the lack of evidence that chimpanzees possess anything remotely resembling a ‘theory of mind. *Philosophical Transactions of the Royal Society, B* 362: 731–744.
- Penn, D.C., K.J. Holyoak, and D.J. Povinelli. 2008. Darwin’s mistake: Explaining the discontinuity between human and nonhuman minds. *The Behavioral and Brain Sciences* 31: 109–130.
- Perner, J., and T. Ruffman. 2005. Infants’ insight into the mind: How deep? *Science* 308: 214–216.
- Pinker. 1994. *The language instinct*. New York: Harper Collins.
- Povinelli, D.J. 2000. *Folk physics for apes: The chimpanzee’s theory of how the world works*. Oxford: Oxford University Press.
- Povinelli, D.J. 2012. *World without weight: Perspectives on an alien mind*. Oxford: Oxford University Press.
- Povinelli, D.J., and J. Vonk. 2003. Chimpanzee minds: Suspiciously human? *Trends in Cognitive Sciences* 7(4): 157–160.
- Povinelli, D.J., and J. Vonk. 2004. We don’t need a microscope to explore the chimpanzee’s mind. *Mind and Language* 19: 1–28.
- Premack, D., and G. Woodruff. 1978. Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences* 1: 515–526.
- Reddy, V. 2008. *How infants know minds*. Cambridge: Harvard University Press.
- Ruffman, T., and J. Perner. 2005. Do infants really understand false belief? Response to Leslie. *Trends in Cognitive Sciences* 9(10): 462–63.
- Senju, A., M.H. Johnson, and G. Csibra. 2006. The development and neural basis of referential gaze perception. *Social Neuroscience* 1(3–4): 220–234.
- Senju, A., V. Southgate, C. Snape, M. Leonard, and G. Csibra. 2011. Do 18-Month-Olds Really Attribute Mental States to Others? : A Critical Test. *Psychological Science* published online 3 June 2011. doi:10.1177/0956797611411584.
- Song, H.-J., K.H. Onishi, R. Baillargeon, and C. Fisher. 2008. Can an agent’s false belief be corrected by an appropriate communication? Psychological reasoning in 18-month-old infants. *Cognition* 109(3): 295–315.
- Southgate, V., A. Senju, and G. Csibra. 2007. Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science* 18(7): 587–592.
- Southgate, V., C. Chevallier, and G. Csibra. 2010. Seventeen-month-olds appeal to false beliefs to interpret others’ referential communication. *Developmental Science* 13(6): 907–912.
- Stack, J., and C. Lewis. 2008. Steering towards a developmental account of infant social understanding. *Human Development* 51: 229–234.
- Surian, L., S. Caldi, and D. Sperber. 2007. Attribution of beliefs by 13-month-old infants. *Psychological Science* 18(7): 580–586.
- Thompson, E. 2007. *Mind in life: Biology, phenomenology and the sciences of mind*. Cambridge: Harvard University Press.
- Tomasello, M., and H. Rakoczy. 2003. What makes human cognition unique? From individual to shared to collective intentionality. *Mind and Language* 18(2): 121–147.
- Tomasello, M., J. Call, and B. Hare. 2003. Chimpanzees understand psychological states – the question is which ones and to what extent. *Trends in Cognitive Sciences* 7: 153–156.
- Tomasello, M., M. Carpenter, J. Call, T. Behne, and H. Moll. 2005. Understanding and sharing intentions: The origins of cultural cognition. *The Behavioral and Brain Sciences* 28: 675–691.
- Träuble, B., V. Marinović, and S. Pauen. 2010. Early theory of mind competencies: Do infants understand others’ beliefs? *Infancy* 15(4): 434–444.
- Trevarthen, C.B. 1979. Communication and cooperation in early infancy: A description of primary intersubjectivity. In *Before speech*, ed. M. Bullowa, 321–48. Cambridge: Cambridge University Press.
- Trevarthen, C., and P. Hubley. 1978. Secondary intersubjectivity: Confidence, confiding and acts of meaning in the first year. In *Action, gesture and symbol: The emergence of language*, ed. A. Lock, 183–229. London: Academic.
- Varela, F., E. Thompson, and E. Rosch. 1991. *The embodied mind*. Cambridge: MIT.
- Vonk, J., and D.J. Povinelli. 2011. Social and physical reasoning in human-reared chimpanzees: new data from a set of preliminary studies. In *Perception, causation, and objectivity. Issues in philosophy and psychology*, ed. N. Eilan, H. Lerman, and J. Roessler. Oxford: Oxford University Press.
- Wellman, H.M., D. Cross, and J. Watson. 2001. A meta-analysis of theory of mind development: The truth about false belief. *Child Development* 72: 655–684.
- Woodward, A.L., and J.A. Sommerville. 2000. Twelve-month-old infants interpret action in context. *Psychological Science* 11: 73–77.