

(In Press) Cultural learning and the reliability of the intentional stance
In: *Content and Consciousness 2.0: Four Decades After (Studies in Brain and Mind Series: Springer)*. Edited by Carlos Munoz-Suárez, with a response by Daniel Dennett.

THE INTENTIONAL STANCE AND CULTURAL LEARNING: A DEVELOPMENTAL FEEDBACK LOOP

John Michael

Department of Cognitive Science
Central European University, Budapest
michaelj@ceu.edu

Abstract: In this paper, I propose a developmental explanation of the reliability of the intentional stance as an interpretive strategy, and by doing so counter an objection to Dennett's intentional stance theory (i.e. the 'If it isn't true, why does it work?' objection). Specifically, young children's use of the intentional stance enables them to learn from and thereby to become more similar to the adults in their culture. As a result, they themselves become increasingly intelligible to other people taking the intentional stance. Thus, the intentional stance and cultural learning constitute a feedback loop that (partially) explains the reliability of the intentional stance, and does so – contra Dennett's realist critics – without appealing to a realist interpretation of the descriptions speakers attach to intentional terms. However, I also suggest that this developmental perspective provides grist to the mill for a *causal* realist interpretation of the reference of intentional terms, insofar the causal interaction between intentional interpretations of behavior and cognitive development provides an anchor that links intentional terms to functional and/or neural processes. Importantly, causal (as opposed to descriptive) theories of reference make it possible to argue that intentional discourse can be referentially anchored to the causal machinery that produces behavior without generating true descriptions of it. I conclude by drawing out some consequences of the developmental perspective for the way in which we conceptualize the assumption of rationality that is at the core of the intentional stance theory.

Keywords: intentional stance, cultural learning, rationality, intentional realism, instrumentalism, causal theory of reference, descriptive theory of reference

1. Introduction

In this paper, I propose a developmental explanation of the reliability of the intentional stance as an interpretive strategy, and by doing so I counter an objection to Dennett's intentional stance theory. I will begin (sec. 2) by briefly recapitulating Dennett's theory, as well as one of the central objections that has animated critical discussions of it, namely that its instrumentalist character prevents it from accounting for the reliability (perhaps even indispensability) of the intentional stance as an interpretive strategy (Richardson 1981; Bechtel 1985; Fodor 1985; Millikan 1993). (Fodor 1985 refers to this as "the 'If it isn't true, why does it work?' problem.") The bulk of the paper will then be devoted to articulating my own proposal.

The core of my proposal is the idea that young children's use of the intentional stance during cognitive development enables them to learn from and thereby become more similar to the adults in their culture, whereby they themselves become increasingly predictable and intelligible for other people taking the intentional stance. Thus, the intentional stance and cultural learning constitute a feedback loop that (partially) explains the reliability of the intentional stance. This proposal commits me to the following three claims, which I will present and defend in turn:

- (i) Children take the intentional stance from early infancy (section 3);
- (ii) Doing so enables cultural learning (section 4);
- (iii) Cultural learning (partially) explains the reliability of the intentional stance (section 5).

After presenting and defending these claims, I will then (section 6) consider the implications of this developmental perspective for the reference of intentional terms. It will become apparent that the proposed feedback loop helps to meet the "If it isn't true, why does it work?" objection and thus to defend Dennett against his realist critics, i.e. it helps to explain why the intentional stance is an effective interpretive strategy without maintaining that adequate intentional explanations of behavior must be true statements about the causal machinery that produces behavior.

It is important to state at the outset that this is not intended as a *general* defense of Dennett's theory or as an argument against realism; it simply rebuts an objection that would otherwise threaten to undermine Dennett's theory. Thus, the account is compatible with various realist positions. In fact, I will be suggesting that the account opens the door to a *causal* realist interpretation of the reference of intentional terms, since the causal interaction between intentional interpretations of behavior and cognitive development provides an anchor that links intentional terms to functional and/or neural processes. Importantly, causal (as opposed to descriptive) theories of reference make it possible to argue that intentional discourse can be referentially anchored to the causal machinery that produces behavior without generating true descriptions of it. Thus, a causal realist interpretation would enable one to resist the strong realist inference that Dennett wants to resist while doing justice to the realist intuition that intentional discourse must be somehow anchored to the functional and/or neural processes underlying behavior.

I will close (section 7) with some reflections on how the developmental account being offered here relates to Dennett's own evolutionary response to the "If it isn't true, why does it work?" objection, and to how it bears upon the assumption of rationality at the core of the intentional stance.

2. *The Intentional Stance*

The core of Dennett's intentional stance theory is the proposal that our everyday practices of explaining and predicting other people's behavior are best characterized in terms of what he calls the *intentional stance*. Taking the intentional stance toward a system (such as another person) is to approach it as an entity "whose behavior can be predicted by the method of attributing beliefs, desires, and rational acumen" (1987: 49). More specifically, an interpreter takes the intentional stance not just by ascribing any old beliefs and desires but by assuming that the system has the beliefs and desires it *ought* to have, and that it reasons rationally from *these* beliefs and desires (Dennett 2008, 1987)¹. In identifying the appropriate beliefs and desires, and working out their effects upon behavior, Dennett proposes that interpreters are guided by the following "rough-and-ready" principles:

- (i) A system will have the beliefs it ought to have, given its perceptual capacities, epistemic needs and biography;
- (ii) A system will have the desires it ought to have, given its biological needs and the most practicable means of satisfying them;
- (iii) A system will perform the actions that it would be rational to perform, given its beliefs and desires (1987: 49).

There are a few key features of the intentional stance that are worth emphasizing. First of all, it is *normative* insofar as it accords a central role to a regulative ideal of rationality, in light of which interpreters try to make sense of target agents' behavior, i.e. they aim to construe it as maximally rational. Secondly, it is *holistic*: the various ascriptions mutually constrain each other in order to retain overall consistency. Thirdly, it is an *idealizing* method, since it assumes rationality as a regulative ideal, although we are of course not really perfectly rational creatures. Rather, we approximate that ideal well enough for the method to be useful (more on this in a moment).

For these reasons², Dennett does not think that interpreting a target agent's behavior on the basis of the intentional stance entails postulating causally salient structures inside the head. In other words, interpretations do not generally aim to pick out the internal physical workings of the system that causally bring about the behavior: "...the beliefs and desires that it (folk psychology) attributes are not – or need not be – presumed to be intervening states of an internal behavior-causing system" (1987: 52). Rather, for an agent to have a particular belief is merely for the attribution of this belief to be compelling to an interpreter, where an interpreter has a characteristic viewpoint and set of goals.

Does this mean that intentional ascriptions are purely relative, that intentional terms do not refer to anything objective, and/or that beliefs and desires are not real? Dennett has consistently resisted such radical consequences, and has therefore tried over the years to carve out an ontological middle ground between realism and eliminativism (sometimes calling it "instrumentalism"³ and sometimes "mild realism"⁴). In fleshing out the sense in which his

¹ Dennett distinguishes various stances that one can take in interpreting the behavior of different systems (Dennett 1978: 237-238; 1971, 1987). Apart from the intentional stance, one can, for example, take the *physical stance*, and assume that a system will be predictable on the basis of physical laws. Or one can take the *design stance*, and treat the system as the product of evolutionary or human design, in which case one will interpret the workings of its parts in terms of their likely functions.

² There are also some other reasons, to discuss which would be beyond the scope of this paper. See, for example, Bechtel 1985 for a discussion of them.

³ E.g. in "Three Kinds of Intentional Psychology", Dennett 1987: 53.

position is instrumentalist, Dennett has drawn upon the distinction between *abstracta* (“calculation-bound entities or logical constructs”) and *illata* (“posited theoretical entities”) (1987: 53), and suggested that beliefs and desires are best understood as *abstracta*. Thus, they are not like dark matter, the existence of which we take to be probable but uncertain because of various observations and theoretical considerations. Rather, they are more like centers of gravity, the existence of which is not a matter of probability but of convention. The question arises, then, if intentional states are abstract objects, just what they are abstractions *from*. The two obvious candidates are brain states and processes, on the one hand, and behavior on the other. Dennett himself favors the latter. In fact, he explicitly contrasts his view with the former view, which he attributes to Fodor, namely that intentional terms pick out “a pattern of structures in the brain,” (ibid: 191). On Dennett’s view, intentional concepts pick out “real patterns” that are “discernible in agents’ (observable) behavior” (2008: 191). The proof of their reality is that recognizing them makes it possible to formulate generalizations and predictions that one could not otherwise formulate.

Dennett has been careful to emphasize, however, that he is not an instrumentalist about psychological or neural states *in general*. Thus, he speaks of his “realism about brains and their various neurophysiological parts, states, and processes” (Dennett 1987: 72), and confirms his agreement that “it is reasonable to consider sensory experiences to be real states of the brain, states whose neurobiological properties will be discovered as cognitive neuroscience proceeds” (Dennett 1993: 210). In fact, he describes himself as being “as staunch a realist as anyone about those core information-storing elements of the brain, whatever they turn out to be, to which our intentional interpretations are anchored” (1987: 70). He just does not expect that those elements will turn out to be “recognizable as the beliefs we purport to distinguish in folk psychology” (1987: 71).

It seems fair to ask, then, just how our intentional concepts are “anchored to” those information-storing elements of the brain that cause behavior if intentional states are abstractions from behavior. And, indeed, realists have queried whether it is possible to account for the reliability (perhaps even indispensability) of the intentional stance as an interpretive strategy without endorsing a more robust realism about the intentional states and rational thought processes that it postulates. In other words, the empirical success of predictions based upon intentional state ascriptions is mysterious if intentional states do not really cause behavior. Fodor (1985) refers to this as “the ‘If it isn’t true, why does it work?’ problem.” (Cf. also Bechtel 1985; Millikan 1993; Dretske 1985).

Although Dennett resists the realist argument that recognizing patterns in behavior would only be useful if they corresponded isomorphically to a second set of patterns within the brain (2008: 201), his talk of “anchoring” does appear to acknowledge that, in order to be useful in explaining and predicting behavior, intentional terms must *somehow* be related to the brain states that cause behavior. In spelling out this relationship, Dennett appeals to evolution, arguing that *evolution is likely to have shaped us in such a way that we approximate the rational agents that the intentional stance posits*. Note that this proposal does not entail a commitment to an isomorphic relation between the concepts and distinctions that structure intentional discourse and the functional or neural levels of description. The idea is that evolution will have selected for *any* functional and/or neural mechanisms that lead to approximately rational behavior, and since there are in principle lots of different mechanisms that could achieve this, which ones actually underlie a particular pattern of behavior will depend on the details of evolutionary history. Thus, it would be foolish to expect to be able to extrapolate this in a straightforward fashion from descriptions couched in intentional terms.

⁴ E.g. in “True Believers”, Dennett 1987: 28, and in “Instrumentalism Reconsidered” Dennett 1987: 71.

However, although various functional and/or neural processes could underlie a pattern of behavior that is nevertheless one and the same pattern as described in intentional terms, one may be inclined to think that those functional and/or neural processes must produce some of the same effects in order to instantiate the same pattern. Take, for example, the following pattern: Jim *sees* an object O being placed at location L, he *hears* it being placed there, he is *told* that it is there, he acquires good reasons to *infer* that it is there, etc. We see a pattern in these cases, namely that Jim acquires the belief that O is at L. Surely, the neural processes giving rise to this belief are different in these different cases, and some of the functional properties will also be different. *But some of the functional properties will also be the same:* Jim will be led to infer that O is not at some other location, to desire to go to L if he desires O, etc. So it seems reasonable to expect that in Jim's brain, the information he attains in these different cases is going to be treated as equivalent regardless of its source, and that this is why it is fruitful to treat these cases as constituting a pattern.

This, at any rate, is the realist intuition. I will be attempting later on show that there is a way of doing more justice to this intuition than Dennett has so far done while stopping short of the standard realist inference that adequate intentional explanations need to be isomorphic with adequate functional or neural explanations. For now, however, let us round out the discussion of the intentional stance theory by briefly considering three sources of concern other than the "If it isn't true, why does it work?" objection:

1. *What about aberrant beliefs and desires?* The first concern is that there is some question as to how central the assumption of rationality is in our interpretations of others' behavior, given that we routinely ascribe beliefs and desires to people that depart from ideal rationality and therefore must appear aberrant from the perspective of the intentional stance theory. We might distinguish three categories of such beliefs and desires. First, there are beliefs and desires that agents ought not have given their perceptual access, epistemic needs and biography, or their biological needs (cf. the three rough-and-ready principles referred to above as constituting the core of the intentional stance). Secondly, there are beliefs and desires that are formed through faulty inferences. Thirdly, there are beliefs and desires that bear utterly non-rational connections to behavior. An example from Stich (1981) illustrates the first category: we may ascribe to Sam the desire for a chocolate bar even if we have already ascribed to him the desire to stay healthy and the belief that he (himself) has a nasty allergy to chocolate. Although it is not irrational on the part of Sam to have this desire (in particular if he actively resists the urge to act upon it), it is a desire that he ought not have given his biological needs, and is therefore in conflict with Dennett's second "rough-and-ready principle" (see above). One might therefore say, tongue-in-cheek, that Mother Nature was irrational in endowing him with this desire. Secondly, even when they start out from perfectly good beliefs and desires, people sometimes make less-than-ideal inferences. For example, most people have the intuition that a politically active young woman with a college degree and feminist political views is more likely to be a feminist and a bank teller than just a bank teller (Stich and Nichols 2003:145-7; Tversky & Kahnemann 1983; Cf. Nisbett and Ross 1980). Moreover, when people are guided by their emotions, they tend to make all manner of dubious inferences (think of Othello). The problem, of course, is that such failures of rationality are common enough to be predictable and even understandable. Thirdly, we also routinely generate perfectly good explanations of others' behavior by ascribing beliefs, desires and other intentional states on the basis of their utterly *non-rational* effects upon behavior (cf. Goldman 2006, chap. 3). For example, the poker player's twitch reveals that she is bluffing, the young man's blushing reveals that he is embarrassed or in love, etc.

2. *Can evolution underwrite an assumption of ideal rationality?* A second issue – also raised by Stich (1981) – is that then even if the assumption of ideal rationality were empirically

adequate, evolution may not explain that adequacy. For evolution does not select for true beliefs but merely useful ones. Sometimes – perhaps often – false beliefs may be more useful than true/rational ones. To take an example from Stich (1981): if organisms in an environment with an abundance of food discover that some particular yellow fruit is poisonous, they may do quite well with a strategy of assuming all yellow fruit to be poisonous and avoiding it (“better safe than sorry”). Whatever one thinks of this example, it does seem plausible that evolution should favor some useful but false beliefs (and some useful but rationally sub-optimal belief-forming processes).

3. *Rationality underdetermines ascriptions.* A third source of concern is that the assumption of rationality at the core of the intentional stance theory – irrespective of its accuracy and its justification – underdetermines most everyday intentional ascriptions. Consider neutral beliefs and desires: Apart from beliefs and desires that one ought not have, we also routinely ascribe beliefs and desires that are *rationally neutral* – that is, there is no reason why one ought to have them. We ascribe to young children, for example, the belief that the Easter Bunny has hidden some eggs in the garden. Similarly, we routinely ascribe neutral desires to people, such as the desire to watch television, or the desire to drink some coffee. Moreover, there is always the problem of figuring out which beliefs and desires people ought to have, given that they desire what it makes evolutionary sense to desire. For example, in one particular culture, you should desire shells, because you can use shells to acquire food and other useful things, whereas in another culture you should desire money, because money is what is used to acquire food, etc. Thus, cultural knowledge is required in order to reach a level of specification that is useful for predicting/explaining/influencing people’s behavior in everyday life.

In fairness, it must be noted that Dennett has in fact always acknowledged these kinds of limitations, and avowed that the intentional stance is supplemented and sometimes corrected with some empirical generalizations that people learn inductively (1987: 54). After all, he does not claim that the intentional stance theory captures all of folk psychology, but only the core, rationalizing part of it. Thus, it is really no problem for him to accept that people also sometimes explain and predict behavior by ascribing aberrant beliefs and desires⁵ or to non-rational links between intentional states and behavior. The under-determination objection is perhaps trickier, insofar as it seems to imply that explanations and predictions of others’ behavior *always* depend (in part) upon cultural knowledge that has nothing to do with rationality. Thus, cultural knowledge is not just an additional interpretive tool that complements the intentional stance but is required in order to apply the intentional stance.

None of these critical observations constitutes a knock-down objection, nor indeed does the “If it isn’t true, why does it work?” objection. In raising them, my intention is to set out some of the unresolved issues that it would be desirable for an account such as mine to resolve. And I will be trying to show later on (sections 6 and 7) that the developmental perspective outlined here (sections 3-5) generates novel and satisfying responses to these objections in a way that is largely compatible with Dennett’s theory.

⁵ This applies not only to beliefs and desires that are formed through faulty inferences but also beliefs and desires that agents ought not have, given their perceptual access, epistemic needs, biography and biological needs, and to beliefs and desires that have non-rational connections to behavior.

3. *The Intentional Stance in Early Infancy*

The first step in my argument will be to review some evidence that children take the intentional stance from early infancy. One potential obstacle to establishing that infants take the intentional stance sufficiently early for it to enable cultural learning is that children do not generally pass explicit verbal false belief tests until they are over four years of age (Wimmer and Perner 1983; Griffin and Baron-Cohen 2002; Apperly 2011) – if children do not pass this litmus test for theory of mind, or mindreading, it may seem unlikely that they ascribe beliefs, desires and rational thought processes in early childhood.

But taking the intentional stance does not require infants to *conceptualize* or to *explicitly* ascribe mental states. Rather, it only requires that the expectations they form about other agents' behavior reflect a sensitivity to those agents' goals, their strategies for attaining those goals, and/or basic mental states such as attentional states and emotions. And there is a wealth of research in developmental psychology suggesting that this is the case. By six months, infants' gaze following reveals a sensitivity to attentional states (Senju and Csibra, 2008). In fact, Reddy has argued persuasively that two-month-olds respond to others' attention in ways that suggest that they experience others as attentional beings (e.g. Reddy 2003). By around six months, the phenomenon of affect attunement attests to a sensitivity to others' emotions (Stern 1985); by 6.5 months infants perceive goal-related movements on the part of geometric shapes (Gergely and Csibra 2003); by nine months, they distinguish cases where an agent does not do something because she is not trying from cases when she is trying but unable (Behne et al. 2005); by ten months, they parse streams of behavior into units that correspond to what adults would see as separate actions (Baldwin et al. 2001).

Moreover, although the claim that infants take the intentional stance does not imply that they use the concept of belief or other mental states to understand others' behavior, some theorists argue that there is such evidence (Baillargeon et al. 2010; Carey 2009), and I think that a strong case can be made that they at least *partially* master such concepts by the end of the first year, and that the case begins to get quite strong by around 18 months. At nine months, for example, children already grasp something about the relations among beliefs and desires, as evinced by their expectation that agents will be happy when a goal is achieved and disappointed when the goal is not reached (Tomasello et al. 2005:6). In a study involving 15-month-olds, Träuble et al. (2010) used an apparatus designed such that an agent could cause a ball to be transferred from one bucket to another by manipulating the apparatus without seeing it (i.e. with her back turned). The finding was that infants expect an agent not to have a false belief even though she did not see the object transfer because she was turned the other way. This demonstrates an impressive ability to reason flexibly about the effects that various kinds of evidence (even non-perceptual evidence) will have on agents' beliefs. In other words, the infants must recognize a pattern insofar as they must interpret the adult agent's manipulation of the apparatus as being relevantly similar to (and thus constituting a pattern with) the agent's visual perception of the location of the ball. Similarly, Song, Onishi & Baillargeon (2008) found that 18-month-old infants' expectations are modulated if the experimenter *communicates* to the agent that the ball has been moved but not if she says merely that she likes the ball.

As already noted, taking the intentional stance does not presuppose this level of sophistication. In fact, Zawidzki reserves the term "intentional stance" for non-mentalistic interpretation. In characterizing the intentional stance, he writes: "Such a framework is not meant as a model of psychological processes; it is a framework for interpreting *behavior*, not mindreading" (Zawidzki, 2013: 38). Although I think it is far from clear that Dennett really wants to restrict the term in this way, I will not dispute this, since Dennett exegesis is beside the point here, and Zawidzki is free to restrict the term in this way irrespective of whether Dennett would approve. But I will not adopt the proposed restriction: it seems to me that the

intentional stance theory can be interpreted as an *analysis* of sophisticated mental concepts rather than an alternative to them, and that is the interpretation I will be working with here.

4. *The Intentional Stance and Cultural Learning*

The next step is to establish that taking the intentional stance enables cultural learning. Tomasello, Kruger and Ratner (1993) distinguish three types of cultural learning that are either unique to humans or at least far more pronounced in humans than in any other animals, and which, crucially, depend upon learners and teachers understanding each other as beings who ‘have intentional and mental lives like their own’ (Tomasello 1999, 7): imitative learning, collaborative learning, and instructed learning.

Consider imitation. In Tomasello’s (Tomasello 1999, Tomasello et al. 2005) terminology, what distinguishes imitation from emulation is that the learner focuses not only on the environmental effect of an observed action but on the observed agent’s goals and strategies. This allows the learner to understand the agent as rationally selecting an appropriate sequence of actions to realize a goal. From about 18 months, infants tend to imitate incomplete but intended actions rather than replicating the exact behavior they have seen, e.g. when an agent tries but fails to close a drawer (Meltzoff 1995). Also, around 14 months, they selectively imitate features of an action that are relevant to the goal of the action – unless the manner in which the agent performed the action appears irrational given the goal, in which case they imitate the particular manner. Thus, for example, if an agent uses her head to turn on a light because her hands are occupied, the child will use his hands to turn on the light, presumably understanding that the adult rationally chose to use her head only because her hands were unavailable. If, however, the adult uses her head even though her hands are free, the child will tend to use his head too (Gergely et al. 2002). The interpretation offered by the authors of this study is that in the first condition (with the hands occupied), the child thinks that the adult only used her head because her hands were occupied, i.e. that she would otherwise have used her hands, given that using the hands would have been the most efficient strategy. In the latter case, however (with the hands free), the child can discern no such reason why the adult used her head, and he therefore reverts to a default assumption that the adult is teaching him something new, i.e. something that he does not understand yet. In sum, the authors maintain that the children in the study were interpreting the agent’s behavior in terms of goals and rational strategies to attain those goals, and that their imitative learning was guided by this interpretation.

Apart from acquainting children with new activities and objects that are common in their culture, imitation also shapes their development in numerous subtle ways. One aspect of moral development, for example, is the acquisition of appropriate behaviors for consoling others in distress, and there is evidence that young children tend to imitate the consolatory behaviors that have brought relief to them in the past when confronted with other individuals in distress, such as offering soothing physical contact or presenting objects that provide comfort or distraction (Hoffman 2000). It is true that this sort of imitation may not always require an understanding of the intention to console or of the emotional state of the person to be consoled. However, it is telling that around 18 months, when most children are able to make a distinction between self and other, as evinced by their ability to recognize themselves in a mirror (Lewis and Brooks-Gunn 1979), they also begin to react with empathic and sympathetic responses to victims of distress, and with appropriate, other-directed comforting and prosocial behavior (Bischof-Köhler 1991; Zahn-Waxler, Radke-Yarrow, Wagner, & Chapman 1992; Eisenberg et al. 2006; Vaish et al. 2009). This strongly suggests that infants in the second year of life not only react to others’ emotions, as might occur in the case of emotional contagion, but do so in a way that reflects a sensitivity to the fact that some specific

other person is experiencing some specific kind of emotion. More generally, an understanding of moral norms depends upon detection of others' mental states (especially emotions) in order to recognize when an action has caused distress to others, when adults have disapproved of actions, and when another individual is experiencing a negative emotion and is thus a candidate for consolation (Cf. Prinz 2005).

To turn to a different kind of cultural learning, namely instructional learning, there is also evidence that the intentional stance plays a crucial role here. In particular Csibra and Gergely (2012) argue that a "teleological" stance⁶ enables children to interpret pedagogical cues from adults as indicating an intention to instruct them⁷. Thus, eye contact and other ostensive signals cause young children to expect to be imparted with generic shared knowledge and to adopt appropriate learning strategies. For example, eye contact leads infants to pay preferential attention to generalizable kind-relevant features of objects that adults refer to (Futo et al 2010; Yoon et al 2008), and to learn actions that incorporate causally opaque means (Gergely et al. 2002).

One striking demonstration of the effects of ostensive signals on learning strategies was provided by Topal and colleagues (2008), who propose that children's perseverative search errors in the A-not-B task may be due to a response on their part to ostensive signals made by the adult experimenter – specifically, eye contact may elicit a "pedagogical learning stance" (Topal et al. 2008: 1832), which leads the infants to expect that the adult intends to teach them some generalizable information, such as that the object is generally located in location A, or that one generally searches for the object in location A. This heuristic then distracts the children from making use of the evidence they have just seen that the object has been moved to location B, and thus leads them to make the perseverative error. And indeed, in a version of the task that does not involve the experimenter attending directly to the child, the perseverative search errors were dramatically reduced.

Understanding others as intentional agents with goals and strategies is also necessary for children in learning how to use tools and symbols, since they must understand to what end the tool or symbol is used. Thus, there has been a great deal of research documenting how children's understanding of adults' attentional states and intentions is crucial for language acquisition. For example, if an adult announces her intention to 'find the toma' and then searches in a number of locations, scowling upon seeing some objects and smiling upon seeing one object, children will learn the new word 'toma' for the object the adult smiles at (Tomasello and Barton 1994). In fact, Southgate et al. (2010) found that 17 month-olds learned to apply a novel word ('sefu') to a toy that an adult falsely believed was hidden in a box if the adult pointed at that box and pronounced the word (after being out of the room while the toy was moved from the box to a different location) (for a review of several similar studies, see Tomasello 1999: 114-116). Moreover, as Tomasello (1999) has emphasized, language acquisition gets going around 12 months, at which time children engage in triadic interactions with an adult and an object, and exhibit pointing behavior to inform others of events they do not know about or to share an attitude about mutually attended events others already know about (Liszkowski et al 2007).

Apart from itself being a cultural artifact and being acquired in part by cultural learning, language enables a multitude of further effects upon cognitive development – "from exposing children to factual information to transforming the way they understand and cognitively represent the world by providing them with multiple, sometimes conflicting, perspective upon phenomena" (Tomasello 1999:163). In acquiring a natural language, children learn to

⁶ Which they explicitly associate with Dennett's intentional stance (Gergely and Csibra 2003).

⁷ The flip-side of this is that adults must take the intentional stance toward infants in order to treat them as candidates for learning. I will return to this point later (see Mameli 2001 and McGeer 2007 for thorough discussions).

partition the world into objects and events in a way specific to their culture, and to categorize the objects and events so partitioned, and to take different perspectives upon them. Thus, one object can be described as “the dog”, “fido”, “the dog over there”, “the golden retriever”, etc., and one event can be described as “The dog bit the man”, “The man was bitten by the dog”, “Fido bit Daddy”, etc. Which of these descriptions is appropriate depends upon the speaker’s communicative goals and upon her evaluation of the listener’s interests, knowledge, etc. The ability to switch among these perspectives and to deploy them flexibly is entrenched through early experiences of disagreeing with others who take different perspectives and in reformulating utterances that have not been understood. Language also enables children to internalize rules, to memorize information and procedures, to talk about their own reasoning processes and other experiences, and to re-describe previously implicit procedural knowledge in explicit symbolic terms, thus enabling greater flexibility and systematicity.

Finally, sensitivity to others’ mental states is also of crucial importance in understanding all manner of norms that structure human sociality, since these derive their binding force not from physical facts but from agreement in people’s attitudes about the statuses of entities, the entitlements and obligations they entail, etc. (Cf. Searle 1995; Gilbert 1990). And there is evidence that children as young as two are sensitive to conventional ways of doing things or using objects, and treat these conventions as normatively binding (Rakoczy et al. 2008). The relative importance of spontaneous imitation (Schmidt et al. 2010) and child-directed pedagogical cues (Gergely and Csibra 2006) is currently a matter of controversy, but both processes depend upon children’s understanding an adult’s goal and strategy and adopting it as the right way to perform the action in question (i.e. the way “we” do things).

5. Cultural Learning and the Reliability of the Intentional Stance

The next step is to close the (developmental feedback) loop by making the case that cultural learning (partially) explains the reliability of the intentional stance. One effect of cultural learning is that children become increasingly similar to the adults in their culture. More precisely, they become increasingly similar to the adults around them *as those adults appear to them on the basis of the interpretations they generate by taking the intentional stance*. This effect ensures that the use of the intentional stance during development increases its reliability as an interpretive strategy. For children’s use of the intentional stance will have shaped their own development such that they themselves approximate the intentional agents that they take others to be. And, if so, they will themselves be more easily intelligible for other interpreters taking the intentional stance.

The flip-side of this, as noted briefly above (footnote 7), is that *adults* also take the intentional stance toward young children, and that this also plays a key role in structuring children’s cognitive development, i.e. by setting up expectations for them to fulfill, and by acquainting them with culture-specific objects, practices, narratives, social roles, etc. For example, gender-specific interpretations of infant behavior (such as boys’ cries more often being interpreted as expressions of anger as opposed to sadness) create expectations that children then conform to (Mameli 2001). Insofar as adults’ interpretation of young children as potentially rational intentional agents facilitates children’s enculturation, it also increases the reliability of the intentional stance, and thus becomes, as Mameli (2001) puts it, a “self-fulfilling prophecy”.

And of course this structuring effect⁸ of taking the intentional stance continues into adulthood. Consider, for example, the “Knobe effect”: people are more likely to interpret an action as intentional when there is a morally negative effect than when there is a morally positive effect, which can be interpreted as suggesting that one of the primary functions of intentional ascriptions is to assign blame and thereby regulate behavior (Knobe 2006). Or, more anecdotally, think of how we sanction others for departing from rational or social norms. Thus, as McGeer (2007) puts it, folk psychology is “a *regulative* practice, moulding the way individuals act, think and operate so that they become well-behaved folk-psychological agents: agents that can be well-predicted and explained using both the concepts and the rationalizing narrative structures of folk psychology” (139, emphasis in original).

Moreover, there is good reason to believe that our interpretations of our own behavior and biographies have an influence on our own actions and choices, and are thus also sometimes self-fulfilling prophecies. Some of Gazzaniga’s research with split-brain patients serves as a dramatic illustration of this. One woman, whose right-hemisphere received the instruction that she should get up and leave the room, and who was then presented with a request to her left hemisphere to explain what she was doing (1995, 1393), confabulated that she had gotten up in order to get a soda – and, crucially, she then really did go and get a soda. Thus, to borrow Zawidzki’s gloss on this example: “whether or not our public self-interpretations are justified or true, we actively work to confirm them” (Zawidzki, 2013: 231).

The proposal being put forward here may at first blush appear to favor simulation theory as opposed to theory theory⁹, since the reliability of simulations is clearly contingent upon a similarity between model and target, whereas theories do not need to be similar to whatever they explain. But in fact the proposal is neutral with respect to this dispute, because a mechanism leading to convergence of target agents and others’ interpretations of them would ensure that the narratives, norms and shared understanding of objects and situations that interpreters draw upon match those that structure targets’ behavior. Moreover, consider some potential sources of difficulty in everyday interpretation. Due to the holistic nature of intentional state ascription, any given interpretative act must be sensitive to a myriad of beliefs and desires that mutually constrain each other. And even if all other relevant mental states are taken into account, they will still underdetermine the interpretation, as there is the further problem of deciding which beliefs, desires and perceptions are relevant at the moment.¹⁰ Convergence of target agents and others’ interpretations of them would help to reduce the search space and thereby increase the utility of whatever interpretive method people use, be it theory or simulation (Cf. Zawidzki, 2013, chap 3).

The mechanism proposed here is circular, but not viciously so. It does not require that young children have the intentional states ascribed by interpreters taking the intentional stance, nor that the intentional stance is a reliable strategy when applied to very young

⁸ Mameli (2001) coined the term “mindshaping” to denote this structuring effect of adults’ intentional interpretations of children; Zawidzki (2013, chapter 2) generalizes it to include cases, like imitative learning, where an agent actively converges upon some external model – such as models of other agents, generated by taking the intentional stance toward them.

⁹ Just to recall: According to theory theory, social cognition is enabled by the ascription of unobservable mental states, which are defined in terms of their nomological relations to perceptions, to behavior, and to other mental states (Carruthers 2009; Gopnik 1993; Baron-Cohen 1995). Simulation theory, in contrast, is based on the idea that we generally understand others by “putting ourselves in their shoes” and using our own cognitive systems to model theirs (i.e. to simulate transitions among mental states, from perceptions to beliefs, from beliefs and desires to behavior, etc.) in which case it would (arguably) be superfluous to represent those nomological psychological relations as such (Gordon 1995; Goldman 2006; Heal 1986).

¹⁰ Cf. Dennett (1978: 125-6) on the frame problem.

children. It requires merely that young children *take others to have such states*¹¹. If they do so, and if this provides them with role models to learn from and thereby to become more similar to, then they will develop in such a way that they subsequently become intelligible to others who also take the intentional stance. Moreover, becoming more like the model agents posited by the intentional stance also adds to children's interpretive resources, which, in turn, enable more learning and thereby more similarity, etc. In this sense, the intentional stance and cultural learning constitute a feedback loop.

6. *The Reference of Intentional Terms*

I set up this discussion of a developmental feedback loop as a response to the “If it isn't true, why does it work?” objection, and suggested that it (partially) explains the reliability of the intentional stance without appealing to intentional realism. In this section, I would like to look a bit more closely at the implications of my proposal for the realism/instrumentalism debate – and more generally, for the reference of intentional terms.

To start out by adapting one of Dennett's metaphors, what this developmental perspective encourages us to consider is that intentional states are patterns that we not only recognize and exploit but which we actively contribute to creating and sustaining. Indeed, our recognition of them is part of a causal explanation of how they are created and sustained from one generation to the next. Many patterns have this kind of dynamic structure¹²: when knitting a sweater with a pattern, one presumably monitors (and recognizes) the emerging pattern, and also modulates one's ongoing actions to ensure that the pattern is maintained. Or, if one is playing in a jazz band, one may respond to some perceived pattern in the music by repeating or completing it, whereupon one of the other musicians might do the same, etc.

What I would like to suggest is that this entanglement of pattern recognition, on the one hand, and pattern etiology on the other, provides an additional justification for the belief that those patterns indeed exist, because our recognition of the patterns enables us to further embed them in their respective target systems. Thus, the developmental feedback loop provides us with an additional reason to think that intentional discourse really does latch onto real behavioral patterns in the world.

Importantly, this does not require us to endorse the strong realist claim that those behavioral patterns must be mirrored by a second set of patterns that underlie behavior, i.e. that true descriptions of behavior must be isomorphic to true descriptions of the neural and/or functional processes that cause that behavior. For, although children's brains are surely molded by enculturation – and more specifically, as I am suggesting, by their use of the intentional stance to engage in cultural learning – they are molded *indirectly*. The feedback loop runs between behavioral patterns and interpretations of behavior, causing *these* to converge with each other, not the functional or neural states underlying behavior. The process must of course be enabled by functional and neural changes, but these may be just whatever changes are necessary in order to support cultural learning and thus to bring about the convergence between behavior and interpretations thereof. Thus, in order to explain why the intentional stance is a reliable strategy, it is unnecessary to postulate an isomorphism between behavioral patterns and a second set of underlying patterns that bring about those behavioral patterns.

¹¹ The effect is of course compounded by others taking the intentional stance towards them, as Mameli, McGeer and Zawidzki describe. Again, this does not require adults' intentional interpretations of children be reliable but only that they have a causal influence upon children's development.

¹² Cf. Ian Hacking's reflections on cases in which our classificatory labels (e.g. “multiple personality syndrome”, “homosexuality”) latch onto existing targets but also lead to changes in those targets, generating what he calls “looping effects” (E.g. Hacking 2002).

However, as I noted above (section 2) in discussing the “If it isn’t true, why does it work?” objection, there is a lingering realist intuition that sometimes, when we see various cases as constituting a pattern, we do so by assuming some common underlying mechanism. For example: Jim *sees* an object O being placed at location L, he *hears* it being placed there, he is *told* that it is there, he acquires good reasons to *infer* that it is there, etc. We see a pattern in these cases by seeing them as instances of belief formation. Now I would like to suggest that the developmental account also makes it possible to do justice to this intuition while retaining the core Dennettian dictum that adequate intentional explanations *need not be true of the causal machinery that produces behavior*.

The basic idea is to appeal to causal theories of reference to articulate the notion that the proposed feedback loop creates a causal relationship between intentional discourse and the functional and neural mechanisms underlying behavior, and that this causal relationship provides an anchor that links the reference of intentional terms to the functional and/or neural processes that underlie behavior. Importantly, causal (as opposed to descriptive) theories of reference do not require terms or concepts to be associated with true descriptions of their referents in order to refer to them, because they maintain that reference is fixed by causal interaction with referents, not by true descriptions of them.

As a result, causal theories are better suited than descriptive theories to account for the fact that scientists (and people in general) often make referential connections despite differences in the meanings which they and others attach to terms. Similarly, they are also better than descriptive theories in accounting for cases in the history of science where theories about some entity (e.g. the electron) have undergone dramatic change while scientists have taken themselves nevertheless to be investigating the same thing (but just to have been mistaken in their description of it).¹³

By analogy with scientists who are able to measure the effects of some novel entity and even to manipulate it in a more-or-less controlled manner prior to formulating a theory about its nature, we may think of children as tracking others’ intentional states before they are capable of describing those states. They are, as it were, evolutionarily endowed with “devices” for tracking intentional states, such as eye-gaze following, emotional resonance, and mechanisms that more or less automatically detect and parse intentional action and identify goals. Although very young children have little in the way of explicit descriptions of what these “devices” track, and the descriptions that they subsequently acquire undergo various shifts over the course of cognitive development, these “devices” continue to track the same states into adulthood. Moreover, the various descriptive layers that are added on during development inherit this causal link to others’ mental states and – by virtue of what I have been calling a developmental feedback loop – extend it to ever more fine-grained and more sophisticated mental states.

Thus, causal theories of reference do not require the explanations or descriptions generated by the intentional stance *to be true of* the causal machinery that produces behavior

¹³ It is important to note that pure causal theories are not without problems. For one thing, they make referential continuity too easy (and are thus unable to make sense of cases of referential failure), since all they require is consistency of a causal relation between agent and entity (e.g. Schouten and De Jong 1998). Moreover, there seems to be no theory-free way to pick out a natural kind or a causal entity in the first place, given that any phenomenon instantiates numerous natural kinds and many causal elements acting conjointly. A sample tiger, for instance, instantiates a species, but also a genus, and so on. The lesson seems to be that some level of description is unavoidable both in reference-fixing and reference-transmission across theory-change. Such considerations have informed more recent *causal-descriptive* theories of reference (e.g. Psillos (1999)).

in order for them *to refer to* that machinery. This makes it possible to argue that intentional discourse refers to the real causes of behavior without correctly describing them. Thus, it does justice to the realist intuition that intentional discourse must be anchored to functional and/or neural processes, and yet at the same time it avoids the strong realist inference that adequate intentional explanations of behavior must generate true descriptions of the causal processes underlying behavior. To put this conclusion in terms of Dennett's pattern metaphor: even if the jazz musician is mistaken about the structure of the pattern that she picks up on and repeats (e.g. because she absent-mindedly thinks it consists of fourths instead of fifths), then hears it again and picks it up again etc., this iterated progression gives her confidence that she really is engaged with the pattern.

It may seem strange, after fending off the "If it isn't true, why does it work" objection, to switch back after all to a realist position. However, I submit that by retaining Dennett's commitment to the claim that adequate intentional descriptions of behavior need not be true explanations of the causal processes underlying behavior (i.e. they need not be isomorphic to adequate functional or neural explanations), the causal realist option put on the table here remains compatible with the general spirit of Dennett's theory. I must emphasize, however, that causal realism is not required in order for the developmental account to provide a rebuttal to the "If it isn't true, why does it work" objection. It is simply a further theoretical option for conceptualizing the link between intentional discourse and the functional and /or neural processes underlying behavior.

7. *Evolution, Development and Rationality*

How does the developmental perspective outlined here bear upon Dennett's evolutionary response to the 'If it isn't true, why does it work?' objection? Recall that, in discussing some limitations upon Dennett's response (section 2), I noted that we routinely ascribe beliefs and desires that depart from ideal rationality and therefore must appear aberrant from the perspective of the intentional stance theory. More specifically, we fully expect people sometimes (i) to have beliefs and desires that they ought not have, given their perceptual access, epistemic needs, biography and biological needs, (ii) to draw inferences that are not logically sound, and (iii) to act in ways that are predictable, but not rationally explicable, in light of their beliefs and desires. In discussing these cases, I also noted that Dennett has always acknowledged that the intentional stance must be supplemented and sometimes corrected with the help of empirical generalizations that people learn inductively. So I am not claiming that they undermine his position. However, adopting the developmental perspective enables us to conceptualize these supplementary resources as part and parcel of the intentional stance theory rather than as ad hoc additions: the role that the intentional stance plays in shaping cognitive development helps to account for interpreters' ability to correctly anticipate not only rational behavior but also departures from ideal rationality (as long as those departures are typical within a particular culture).

I also pointed out in section 2 that concrete behavioral predictions are frequently underdetermined by evolutionary considerations, and that cultural knowledge is therefore required in order to reach a level of specification that is useful for predicting/explaining/influencing people's behavior in everyday life. It is rational, for example, to desire resources, but in order to bring this truism to bear in predicting an agent's behavior, it will generally be necessary to know what counts as a resource in their culture (e.g. shells, money, etc.). Note that cultural knowledge in this sense is not merely an additional tool that supplements the assumption of rationality and can be used in some range of cases where the assumption of rationality does not apply; rather, it is required in order to make use of the assumption of rationality *even in those cases where the latter does apply*. In sum, cultural knowledge acquired through cultural learning is necessary in order to specify

what beliefs and desires a target agent should have *given a particular cultural context*, and to fill in gaps that an assumption of ideal rationality does not account for.

The upshot is that although evolutionary considerations constrain the determination of what intentional states to ascribe to others, it is shared developmental history that ultimately enables interpreters to fix the specific contents they ascribe to others, and to do so in ways that are fairly rational but not ideally rational. Indeed, there is likely to be an evolutionary rationale for why the intentional stance should play the shaping role that it does in development. If taking the intentional stance really does enable imitation and other forms of cultural learning, then it may support an inheritance system that has been shaped by the need to transmit behavioral phenotypes reliably from one generation to the next,¹⁴ and by the need to increase homogeneity within human populations in order to facilitate cooperative behavior¹⁵. Thus, evolution may be better suited to underwrite an assumption of culture-specific *imperfect* rationality than one of *ideal* rationality.

8. Conclusions

I have argued that the intentional stance and cultural learning constitute a feedback loop that (partially) explains the reliability of the intentional stance, and does so – contra Dennett's realist critics – without appealing to a realist interpretation of the descriptions speakers attach to intentional terms. I have also suggested that this developmental perspective opens up the possibility of conceptualizing the link between intentional discourse and the functional and/or neural processes underlying behavior in terms of a causal theory of reference: the causal interaction between intentional interpretations of behavior and cognitive development anchors the reference of intentional terms in the functional and/or neural processes underlying behavior – and this anchoring does not require intentional explanations to be true of those functional and/or neural processes.

A further insight generated by the developmental perspective is that it is perhaps not an assumption of ideal rationality that constitutes the core of the intentional stance as an interpretive strategy but an assumption of culture-specific imperfect rationality. Interpreters expect agents to behave just as rationally as people tend to behave in their culture, and to deviate from ideal rationality in ways that are typical within their culture. Moreover, they also make specific predictions that cannot be generated simply by assuming others are (more or less) rational but which also draw upon specific cultural knowledge.

Finally, the developmental perspective outlined here also suggests a slight refinement of the criterion for determining whether to regard a target system as an intentional system. The criterion that Dennett himself has proposed is simply that the intentional stance is likely to be viable if the system in question is the product of natural selection (1978: 8). This criterion provides us with a sensible minimal requirement for applying the intentional stance, but it does not help us to understand why the intentional stance works so much better when applied to other humans, especially humans with similar cultural backgrounds, than when applied to non-human animals, or to plants, bacteria and other cognitively unsophisticated evolved creatures. The developmental perspective, in contrast, enables us to account for these differences by observing that humans are particularly appropriate targets for intentional interpretation because their brains have been shaped by their own use of the intentional stance from infancy onward. This is especially true for humans with similar culture backgrounds, because the intentional stance enables children to learn culture-specific norms and practices. Thus, being the product of natural selection may constitute a minimal criterion for a system to be an apt target for intentional interpretation, but the intentional stance is likely to be

¹⁴ Cf. Shea, 2009.

¹⁵ Cf. Sterelny, 2003; Zawidzki, 2013.

especially useful when applied to a system if it is the case that cultural learning is part of the etiology of the structures that produce that system's behavior.

Acknowledgments

Thanks to Mathieu Arminjon, Yanxia Feng, Ellen Fridland, Rick Griffin and Dan Dennett for stimulating discussions that helped a great deal in preparing this manuscript, and also to Gualtiero Piccinini and to two other (anonymous) reviewers for their constructive criticisms.

References

- Apperly, I. 2011. *Mindreaders: The Cognitive Basis of Theory of Mind*, New York: Psychology Press, 2011.
- Griffin, R. and Baron-Cohen, S. 2002. The Intentional Stance: Developmental and Neurocognitive Perspectives. In (2002) A. Brook & D. Ross (Eds.), *Daniel Dennett: Contemporary Philosophy in Focus*. Cambridge, UK: Cambridge University Press: 83-116
- Apperly I., and S. Butterfill. 2009. Do humans have two systems for tracking beliefs and belief-like states? *Psychological Review* 116 (4): 953-970.
- Baillergeon, R. Scott, R., Z. He. 2010. False-belief understanding in infants. *Trends in Cognitive Sciences* 14 (3): 108-115.
- Baldwin, D. A., Baird, J. A., Saylor, M. M. & Clark, M. A. (2001) Infants parse dynamic action. *Child Development* 72: 708–17.
- Baron-Cohen, S. 1995. *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge, MA: MIT Press.
- Bechtel, W. 1985. Realism, instrumentalism and the intentional stance. *Cognitive Science* 9, 473-497.
- Behne, T., Carpenter, M., Call, J. & Tomasello, M. 2005. Unwilling versus unable: Infants' understanding of intentional action. *Developmental Psychology* 41:328–37.
- Bischof-Köhler, D. 1991. The development of empathy in infants. In M. E. Lamb & H. Keller (Eds.), *Infant development: Perspectives from German speaking countries* (pp. 245-273). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Carruthers, P. 2009. How we know our minds: The relationship between metacognition and mindreading. *Behavioral and Brain Sciences* 32, 121 – 182.
- Carey, S. 2009. *The Origin of Concepts* (1st ed.). Oxford University Press, USA.
- Csibra G., Gergely, G., Szilvia Biro, S., Kiro, O. and M Brockbank. 1999. Goal attribution without agency cues: the perception of 'pure reason' in infancy. *Cognition* 72: 237–267.
- Dennett, D. 2008. Real Patterns. In: Bedau, M., Humphreys, P. (Eds.), *Emergence*, Cambridge, MA: MIT Press (Originally published in: *Journal of Philosophy* 88 (1), 1991: 27-51).
- 1993. Back from the drawing board. In Dahlbom, B. (Ed.) *Dennett and his Critics*. Oxford: Blackwell, 203-235.
- 1987. *The Intentional Stance*. Cambridge, Mass.: MIT Press.
- 1978. *Brainstorms*. Montgomery, VT: Bradford Press.
- 1971. Intentional systems. *Journal of Philosophy*, 68, 87–106.
- 1969. *Content and Consciousness*. London: Routledge.
- Eisenberg, N., & Fabes, R. A. 1998. Prosocial development. In N. Eisenberg (Ed.), *Handbook of child psychology, Vol. 3: Social, emotional, and personality development* (5th ed., pp. 701-778). New York: Wiley & Sons.
- Eisenberg, N., Spinrad, T. L., & Sadovsky, A. 2006. Empathy-related responding in children. In M. Killen & J. G. Smetana (Eds.), *Handbook of moral development* (pp. 517–549).

- Mahwah, NJ: Erlbaum.
- Fodor, J. 1985. Fodor's guide to mental representation: the intelligent auntie's vade-mecum. *Mind*: 76-100.
- Gazzaniga, M. 1995. Consciousness and the cerebral hemispheres. In: *The cognitive neurosciences*, ed. M. Gazzaniga. MIT Press.
- Gergely, G., Bekkering, H. and I. Kiraly. 2002. Rational imitation in preverbal infants. *Nature* 415: 755.
- Gergely, G. and G. Csibra. 2006. Sylvia's recipe: the role of imitation and pedagogy in cultural transmission. In *Roots of Human Sociality: Culture, Cognition and Interaction*. Ed. Enfield, N and S. Levinson, New York: Berg, 229- 255.
- Gergely, G. & Csibra, G. 2003. Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences*, 7(7), 278-292.
- Gilbert, M. 1990. Walking together: a paradigmatic social phenomenon Midwest Studies in Philosophy.
- Goldman, A. 2006. *Simulating minds*. Oxford: Oxford University Press.
- Gopnik, A. 1993. The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences* 16: 1–14.
- Gordon, R. 1995. Simulation without introspection or inference from me to you. In *Mental Simulation: Evaluations and Applications*, ed. T. Stone and M. Davies, 53-67. Oxford: Blackwell.
- Griffin, R. & Baron-Cohen, S. 2002) The intentional stance: developmental and neurocognitive perspectives. In A. Brook & D. Ross (Eds.) *Daniel Dennett*. Cambridge, England: Cambridge University Press, pp. 83-116.
- Hacking, I. 2002. *Historical Ontology*, Cambridge, MA: Harvard University Press
- Heal, J. 1986. Replication and functionalism. In *Language, Mind and Logic*, ed. J. Butterfield, 135-150. Cambridge: Cambridge University Press.
- Hoffman, M. 2000. *Empathy and Moral Development: Implications for Caring and Justice*. New York: Cambridge University Press.
- Knobe, J. 2006. The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology. *Philosophical Studies*. 130: 203-231.
- Lewis, M. & Brooks-Gunn, J. 1979. *Social cognition and the acquisition of self*. New York: Plenum Press. p. 296
- Liszkowski, U., Carpenter, M. and M. Tomasello. 2007. Pointing out new news, old news and absent referents at 12 months of age. *Developmental Science* 10 (2): 1-7.
- Mameli, M. 2001. Mindreading, mindshaping, and evolution. *Biology and Philosophy* 16: 597-628.
- Meltzoff, A.N. 1995. Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31: 838–50.
- Millikan, R. 1993. *White Queen Psychology and Other Essays for Alice*. Cambridge: MIT Press.
- Nichols, S., and S. Stich. 2003. *Mindreading*. Oxford: Oxford University Press.
- Prinz, J. 2005. Imitation and moral development. In *Perspectives on Imitation*. Ed. Hurley, S. and N. Chater. Cambridge: MIT Press: 267-282.
- Psillos, S. 1999. *Scientific Realism: How Science Tracks Truth*, London: Routledge
- Rakoczy, H., Warneken, F. and M. Tomasello. 2008. The Sources of normativity: children's understanding of the normative structure of games. *Developmental Psychology*, 44 (3): 875-881.
- Reddy, V. 2003. On being the object of attention: implications for self–other consciousness. *Trends in Cognitive Sciences*, 7(9), 397-402.
- Schmidt, M. Rakoczy, H. and M. Tomasello. 2010. Young children attribute normativity to novel actions without pedagogy or normative language. *Developmental Science* 14 (3):

- 530-539.
- Schouten, M., De Jong, H. 1998. Defusing eliminative materialism: reference and revision. *Philosophical Explorations* 11 (4): 489-509
- Searle, J. 1995. *The Construction of Social Reality*. New York: Free Press.
- Senju, A., and G. Csibra. 2008. Gaze-following in human infants depends on communicative Signals. *Developmental Science* 18 (9): 678-671.
- Shea, N. 2009. Imitation as an Inheritance System. *Philosophical Transactions of the Royal Society B*, 364: 2429-2443.
- Song, H, Onishi, K., Baillargeon, R., Fisher, C. Can an agent's false belief be corrected through an appropriate communication? Psychological reasoning in 18-month-old infants. *Cognition*, 109 (2008), pp. 295–315
- Southgate, V., Chevallier, C., and Csibra G. 2010. Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science* 13 (6): 917-912.
- Sterelny, K. 2003. *Thought in a Hostile World*. Oxford: Blackwell.
- Stern, D. (1985/1998). *The interpersonal world of the infant: A view from psychoanalysis and developmental psychology*. New York: Basic Books.
- Surian, L., Caldi, S., and D. Sperber. 2007. Attribution of Beliefs by 13-Month-Old Infants. *Psychological Science* 18 (7): 580-586.
- Tomasello, M. 1999. *The Cultural Origins of Human Cognition*. Cambridge: Harvard University Press.
- Tomasello, M. and M. Barton. 1994. Learning words in non-ostensive contexts. *Developmental Psychology* 30: 639-650.
- Tomasello, M. Carpenter, M., Call, J., Behne, T. and H. Moll. 2005. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* 28: 675–735.
- Tomasello, M., Kruger, A., and H. Ratner. 1993. Cultural Learning. *Behavioral and Brain Sciences* 16, 495-552.
- Topal, J., Gergely, G., Miklosi, A., Erdohegyi, A., Csibra, G. 2008. Infants' perseverative search errors are induced by pragmatic misinterpretation. *Science* 321: 1831-3.
- Träuble, B., Marinović, V., & Pauen, S. 2010. Early Theory of Mind Competencies: Do Infants Understand Others' Beliefs? *Infancy*, 15(4), 434-444
- Vaish A, Carpenter M, Tomasello M (2009) Sympathy through affective perspective- taking and its relation to prosocial behavior in toddlers. *Developmental Psychology* 45(2):534–543.
- Wimmer, H., and J. Perner. 1983. Beliefs about beliefs: representation and constraining function of wrong beliefs in children's understanding of deception. *Cognition* 13 (1): 103-128.
- Zahn-Waxler, C., Radke-Yarrow, M., Wagner, E., & Chapman, M. 1992. Development of concern for others. *Developmental Psychology*, 28(1), 126-136.
- Zawidzki, T. 2013. *Mindshaping – A New Framework for Understanding Human Social Cognition*. Cambridge, MA: MIT Press.
- Zawidzki, T. 2008. The function of folk psychology: mind reading or mind shaping? *Philosophical Explorations* 11(3): 193 – 210.