

THE WORLD WE POINT AT

Ashton Smith, M. (submitted). The World We Point At: Two Shared Intentionality Accounts of Declarative Pointing and Uniquely Human Cognition. Submitted to: Special issue of the *Review of Philosophy and Psychology* on “Joint Action: What is Shared?” Editors: Natalie Sebanz & Stephen Butterfill.

Do not cite without explicit permission of the author.

The World We Point At: Two Shared Intentionality Accounts of Declarative Pointing and Uniquely Human Cognition

Dr. Mark Ashton Smith
Department of Psychology
Girne American University
North Cyprus

Contact Information:

Dr. Mark Ashton Smith
Department of Psychology
GAU American University
North Cyprus
m.a.smith.lab@gmail.com

THE WORLD WE POINT AT

Abstract

In this paper I review and critique an influential theoretical framework for explaining uniquely human cognition in the social-pragmatic tradition: the shared intentionality account of Michael Tomasello and his colleagues (Moll & Tomasello 2007; Tomasello & Carpenter 2007; Tomasello 2008). *Shared intentionality*, the central explanatory construct in this framework, is understood primarily in the *joint action* sense of ‘collaborative interactions’ and ‘shared goals and shared action plans for pursuing those goals’. The shared intentionality account of ape and infant pointing provides a useful focus for spelling out the theoretical constructs of the broader framework. After briefly reviewing the literature on pointing in Section I, I explicate the essentials of the shared intentionality account in Section II. In Section III I define three distinct types of ‘intentionality’ - motivational, representational and communicative - each of which needs to play an explanatory role in any account of pointing or uniquely human cognition. I highlight four respects in which the shared intentionality account of infant pointing systematically fails to appreciate the challenges posed by one type of intentionality: the *representational* nature of human cognition – in the sense of its semantic content or aboutness. In Section IV I develop the *Shared Intentionality (R) Hypothesis*. This account of uniquely human cognition, and the declarative pointing that reflects it, builds on the social-pragmatic foundations of Tomasello’s and colleagues’ account but brings to center-stage the shared representational intentionality of human cognition. The *SI(R)* account has interesting implications for *joint action* accounts of cognition more generally.

SECTION I

A question of widespread, interdisciplinary interest is the extent to which pointing gestures shares a common cognitive and motivational infrastructure with fully-fledged adult communication before language has emerged, and to what extent pointing may be ontogenetically and phylogenetically foundational in uniquely human modes of cognition and communication (Gómez 2007; Leavens & Hopkins 1999; Liszkowski 2008, Tomasello 2006). The theoretical significance of pointing gestures is emphasized in social-pragmatic approaches to language acquisition and cognitive development in humans. Social-pragmatic (or ‘usage based’) accounts emphasize the child’s understanding of the intentional actions of others, particularly their communicative intentions (Brunner 1975; Grassmann & Tomasello 2010; Tomasello 2003; Tomasello & Abbot-Smith 2005). In this section, I briefly review the pointing literature. In Section II I review Tomasello and colleagues’ shared intentionality account of ape and infant pointing, and, by extension, their conception of uniquely human cognition. In Section III, after identifying three different senses of ‘intentionality’ (motivational, representational and communicative), I argue that referential/representational intentionality is problematic for the shared intentionality account as it stands. In Section IV I present the *Shared Intentionality (R) Hypothesis* – an account of uniquely human cognition that embraces the social-pragmatic foundations of Tomasello’s and colleagues’ account, but which extends them to account for the shared representational intentionality of human cognition.

Ape Pointing

Apes frequently gesture with limbs and hands. Whereas all primates regularly communicate using vocalizations, orofacial movements and body postures, free brachiomanual gestures, are only typical of humans and apes (Pollick & de Waal 2007). Brachiomanual gestures are used more flexibly to communicate across contexts than facial expressions and vocalizations and they are less closely linked to particular emotions, such as aggression or affiliation, and behavioral and brain imaging studies indicate they are under greater cortical control than facial/vocal signals (Pollick & de Waal 2007; Tagliatela et al. 2008).

Comparative psychologists have discovered a striking behavioral difference between the gesturing behavior of wild apes and humans. “Despite intensive observations of the spontaneous interactions of free-ranging chimpanzees, there is no evidence that these animals ...gesture with the arm and/or index finger toward other objects, animals, or events in space” (Povinelli, Bering & Giambrone

THE WORLD WE POINT AT

2003, p. 41). Wild apes don't *point*. In over a 40 years of intensive field studies, according to Moll and Tomasello, "There has not been a single reliable documentation of any scientist in any part of the world of one ape pointing for another" (Moll & Tomasello 2007 p. 643). By contrast, human infants from around one year of age spontaneously begin to point with the index finger in earnest (Liszkowski et al. 2004; Tomasello 2006). Evidence indicates that pointing is a widespread, if not universal, pattern cross-culturally and a stereotyped, species-typical gesture in humans (Butterworth 2003; Leavens & Hopkins 1999; Povinelli, Bering & Giambrone 2003). Comparative psychologists in a number of labs have argued that wild apes' inability to point for others and its universal presence in infants as young as one year constitutes evidence for a naturally selected behavioral adaptation present in humans but absent in apes (Kita 2003; Leavens, Hopkins & Bard 2005; Povinelli, Bering & Giambrone 2003; Tomasello 1999).

In captivity chimpanzees, bonobos, orangutans, and gorillas, without training commonly point to unreachable food, either with their index finger or more commonly with all fingers extended. 60-70% of all captive chimpanzees point imperatively/requestively to caregivers for out of reach food (Leavens & Hopkins 1999; Leavens, Russell & Hopkins 2005). Most of these pointing incidents are reported to be 'spontaneous', developing in the absence of explicit training (Call & Tomasello 1994). Pointing by apes is communicative and not simply a type of grasping in as far as they generally do not point when an observer is absent or looking away from the pointing individual (Call & Tomasello 1994; Leavens, Hopkins & Bard 1996). Head direction, indicative of engagement of *attention*, is important in ape pointing. In their pointing gestures, nonhuman primates appear to interpret and exploit information about the attentional states of interactants, and there are close similarities between how humans and other apes process information about attention. Humans share with a number of primates the ability to use attention cues involving posture and orientation of the eyes, head and body of others to orient to or gain access to external objects or events. For example, there is a nearly complete overlap in the gaze-following behavior of chimpanzees and 18-month-old human infants (Povinelli 2001). *Intentional communication* has been operationally defined in terms of the following criteria: (M) 'attention-getting' behaviors; (b) gaze alternation between an addressee and a distal object or event of interest, and (c) persistence and flexibility to achieve a goal (Bard 1992; Bates, Camaioni, & Volterra 1975). All of these are commonly reported as accompanying ape pointing, indicating that they intentionally communicate with their pointing (review, Leavens & Hopkins 1999).

Pointing Taxonomies

One interpretation of the biological status of ape pointing is that it develops epigenetically or as an exaption of existing cognitive-communicative skills (Gomez 2007; Leavens, Hopkins & Bard, 2005). Leavens and colleagues (2005) have argued that pointing emerges as apes in captivity face a foraging problem unique to captivity in which they cannot acquire desirable food except through manipulation of a human. Pointing thus develops as an epigenetic consequence of the ecological constraints of captivity. The traditional consensus has been that while ape pointing is used exclusively for instrumentally requesting/directing the behavior of others as ‘social tools’, humans will point to simply to direct and engage joint attention and interest in an object. Apes in captivity use pointing only ‘imperatively’, generally to obtain food (Call & Tomasello 1996; Leavens & Hopkins, 1999; Leavens, Hopkins & Bard 2005). Tomasello and Carpenter (2005) repeatedly used procedures that reliably illicit declarative pointing from young human infants, and they were unable to induce any declarative pointing from any of three young chimpanzees. In notable contrast to human infants, no ape - whether wild or in captivity - has been observed to point for another ape for the sake of jointly engaging attention, interest or affect to some external entity, in the absence of a clear instrumental goal to obtain the object. I shall call this type of pointing *declarative pointing* – a designation that embraces the ‘declarative’ and ‘expressive declarative’ and ‘deictic’ taxa described below.

The first theoretical account of infant pointing, that of Bates, Camaioni, and Volterra (1975), captured this contrast, distinguishing between *proto-imperative* pointing where the gesture is used to get adults to retrieve things for them, and *proto-declarative* pointing where the gesture is used to get the adult to attend to an external entity in order to engage the adult’s attention (review, Tomasello, Carpenter & Liszkowski, 2007). In humans, both types of pointing are exhibited around one year of age (Carpenter, Nagell & Tomasello 1998; Leung & Rheingold 1981). A number of studies have observed that the proportion of gestures that would be traditionally be classified as proto-declarative tends to increase in human samples from younger than 1 year of age to approximately 2 years whereas the proportion of ‘reaches’ or proto-imperatives tends to remain relatively constant (Franco & Butterworth 1996). The different developmental pattern is additional evidence for a different functional basis for the two gestures in infants.

THE WORLD WE POINT AT

Liszkowski and colleagues (Liszkowski et al., 2006) have described another type of uniquely human pointing in 12 month old infants: *informative* pointing. This occurs when the infant's intention is to provide the addressee with information she needs or would be interested in. Tomasello, Carpenter and Liszkowski (2007) classify informative pointing as a subtype of 'declarative' pointing, distinguishing between 'declaratives as informatives' and 'declaratives as expressives' in which the motive is to share an attitude (such as enthusiasm or interest) with an adult about a referent. The motivation for expressive pointing is to share emotions/attitudes on this account, not simply to attract an adult's attention by means of an object as Bates and colleagues' proposed. Proto-imperative pointing is, in addition, proposed to lie on a continuum. Some imperative pointing is based on individualistic motives by inducing the addressee as a 'social-causal tool', while others are based more on cooperation by telling the address what is wanted in more of an 'indirect request' that the adult know her goal and decide to help her (Schwe & Markman, 1997). This is called *requestive pointing*. Recent longitudinal studies provide some support for this taxonomy based on functional, hand shape and laterality patterns (Cochet & Vauclair, 2010).

Intentionally communicative pointing gestures in infants have also been classified as *deictic* (Bates, 1979). Deictic (deixis, Greek "to show") gestures show or present a referent in the immediate spatiotemporal environment, the most prototypical being declarative pointing. Classically, infant gestures such as showing and giving have also been classified as deictic gestures (Bates, 1979). Deictic gestures are used to communicate referentially in adults, and human infants are cognitively and motivationally adapted for referential communication with the pointing gesture (Csibra & Gergely 2009; Liszkowski, 2008). This taxonomic tradition stresses the significance of the referent, rather than the engagement of an adult's attention or emotional response.

Pointing And Joint Attention

Referential communication occurs in joint attentional interactions (review Eilan et al. 2005). Tomasello defines the 'referential triangle' as a joint-attention structure incorporating the "child, adult, and the object or event to which they share attention" (Tomasello, 1999 p. 62). Pointing does not have priority over lexical items in interpreting acts of reference unless it is accompanied by gaze alternation between child and object (Grassman & Tomasello 2010; Jaswal & Hansen 2006). The structure of joint attention in declarative pointing has been investigated directly by Liszkowski and colleagues (2004)

THE WORLD WE POINT AT

who found that if the adult did not jointly attend to the object with the infant (by alternating gaze between infant and object and emoting positively on it)—but instead either (i) just attended to the object without looking to the infant or (ii) only looked and emoted positively to the infant while ignoring the object—the infants were dissatisfied and tried to correct the situation. Thus joint attention has a clear triadic structure for the infant.

SECTION II

The Shared Intentionality Account Of Why Apes Cannot Point Declaratively

In this section I review the explanation offered by Tomasello and his colleagues' socio-pragmatic account of why apes can't point declaratively or informatively. According to their shared intentionality account, apes cannot point declaratively because such pointing requires two codependent socio-pragmatic skills that they lack (Rakoczy & Tomasello 2007; Moll & Tomasello 2007): (1) The ability to understand a cooperative *communicative intention*: that a pointing gesture of another individual is intended to be informative and for their benefit. (2) The ability to participate in a *joint attention frame* or *common ground* that determines the *meaning* of the communicative intention behind the pointing, anchoring the *referent* of the pointing act.

Data from the *object-choice task* has been used in support of this hypothesis (review, Call & Tomasello 2005). In this task a human (the 'hider') hides some food in one of two containers. Then another human ('the helper') tilts the container to show the ape where it is. After this induction, the hider again places the food in one of the containers but now the helper uses a communicative cue to indicate the relevant container by pointing to (and in other variants, gazing at, or tapping on, or placing a marker on top of) the relevant container. The results are the same in all variants of this task: chimpanzees most often choose the container randomly, failing to understand the obvious communicative cue. Those who eventually perform well take dozens of trials to learn. Human infants, by contrast, can understand pointing gestures in an object-choice task where there were clear pragmatic cues to intentional communication spontaneously from as early as 14 months of age (Behne, Carpenter & Tomasello 2005).

According to Tomasello and his colleagues, understanding declarative pointing requires understanding *communicative intentions* which they define as "an intention to inform me of something—which is an intention toward my intentional states (an embedded intention)" (Tomasello 2006, p. 509).

THE WORLD WE POINT AT

...to understand pointing, the subject needs to understand more than the individual goal-directed behavior. She needs to understand that by pointing towards a location, the other attempts to communicate to her where a desired object is located; that the other tries to inform her about something that is relevant for her. So the ape would need to understand something about the directedness towards itself ('this is for me!') and about the communicative intention behind the gesture in order to profit from it. (Moll & Tomasello 2007, p. 644).

On this account there is a critical sense in which finding the food in the object-choice task is a cooperative *joint project* within which communicative intentions make sense: the helper's intentionally informative cue is relevant to the search task the subject is facing, and thus is incorporated into some shared comprehension of 'our project' in the game of 'I show you find'. On this account there is thus a *shared intentionality* – or 'we intentionality' - in which "participants have a shared goal and coordinated action roles for pursuing that shared goal (Bratman 1992).

The joint project aspect of communicative intentions is elaborated in the important theoretical notion of a *joint attentional frame* or *common ground* of communication (Clark & Brennan 1991).

It is ... the joint attentional frame, or common communicative ground, emanating from the joint project which gives the pointing gesture its meaning in specific contexts. (Rakoczy & Tomasello 2007 p. 120)

Rakoczy and Tomasello give the following example of a joint attention frame:

if you encounter me on the street and I simply point to the side of a building, the appropriate response would be 'Huh?' But if we both know together that you are searching for your new dentist's office, then the point is immediately meaningful.

(p. 120)

In this example, by virtue of the common ground the pointing has a *relevance* to the 'searching for the dentist's office' task, which – in the act of pointing – both individuals participate in together in a shared intentional way. The common ground functions to pick out the relevant *referent* of the pointing which both individuals can then jointly attend to. Pointing gestures thus have the following Gricean schema: "she intends that I attend to X (and wants us to know this together) for some reason relevant to our common ground" (Tomasello, Carpenter & Liszkowski 2007, p. 708). Infants of 18

THE WORLD WE POINT AT

months clearly understand that the functioning of a joint attentional frame is specific to those people who share it.

That the meaning of a pointing gesture depends on a shared common ground has been demonstrated by Liebal, Carpenter & Tomasello (2010) in 18 month olds (with mixed results for 14 month olds) who participating in different tasks with an adult. How an adult act of pointing was interpreted was shown to depend critically on the shared activity (common ground). Moreover, Tomasello and Haberl (2003) had an adult say to 12- and 18-month-old infants “Oh, wow! That’s so cool! Can you give it to me?” while gesturing ambiguously in the direction of three objects, two of which the adult and the child had played with recently – thus establishing a common ground - and one which was new to the adult (although not the child). Infants gave the adult the object that was novel. Infants knew which objects the adult had experienced, and which he had not. By contrast, apes in a similar paradigm did not distinguish between what was new and what was old for the human (Tomasello & Carpenter 2005). Other studies suggest that infants as young as 12-14 months of age attend to and register an adult’s experience most readily when they are jointly attending with that person in a joint attention frame (Moll, Carpenter & Tomasello, 2007; Tomasello & Haberl, 2003). Joint attention is understood as a type of object directed ‘inter-subjectivity’ in which psychological states are mutually acknowledged and shared:

Joint attention is not just two individuals looking at the same thing at the same time. Joint attention requires that each of the individuals knows that the other is attending to the same thing as they are attending to; that is what makes it a joint, rather than a merely simultaneous, activity. (Call & Tomasello 2005)

A suite of joint-attentional skills are generally in place by 12-15 months of age, and include, early on, simply registering the adult’s attention and behavior towards outside entities by following an adult’s gaze to objects. Subsequently, infants actively direct adult attention and behavior to objects and events using communicative gestures such as declarative pointing, but also holding up an object to show it to someone (Moore & Dunham 1995; Tomasello 1999). They emerge in close developmental synchrony, with 80% of infants mastering all these abilities within a four month window in one study (Carpenter, Nagell & Tomasello 1998)

Returning to the question at hand, apes do not understand the communicative intention behind the pointing in the object-choice task because they cannot establish a cooperative joint attention frame

THE WORLD WE POINT AT

with the experimenter to determine a relevant referent ('food in *that* bucket'). To understand the relevance of the gesture to their goal of finding the food, they would have to understand that it signals information *for them* from an altruistic donor in the context of a joint project. They are incapable of this kind of shared intentionality. Thus "they follow the point to the bucket and say, in effect, 'A bucket. So what? Now where's the food?'" (Tomasello & Carpenter 2007, p. 122). Shared intentionality has been defined as a type of intentionality "in which participants have a shared goal and coordinated action roles for pursuing that shared goal" (Tomasello, 2006, p. 516). In shared intentional situations, "the goals and intentions of each interactant must include as content something of the goals and intentions of the other" (Ibid p. 516). Apes – unlike human infants – are incapable of shared intentionality.

Beyond Pointing: Collective Intentionality

Common ground communicative intention skills are hypothesized to develop in conjunction with a developing understanding of *differences of perspective* on the *same* external entities, from which emerge uniquely human *perspectival* or *dialogic representations*, where both first and third persons' perspectives are simultaneously incorporated in a single representational format from a 'bird's eye' point of view (Tomasello et al. 2005; Behne, Carpenter & Tomasello, 2005). These representations greatly expand the sphere of shared intentionality and – it is proposed - fundamentally transform human cognition and behavior. Symbolic language is conceived as a "historically evolved inventory of symbolic devices for directing others' attention perspectivally in myriad ways" (Tomasello, Carpenter & Liszkowski 2007, p. 719), and perspectival representations enable the imitative learning of communicative conventions with linguistic symbols as "inherently bidirectional coordination devices with reversible roles" (Tomasello 2006, p. 516). Symbolic communicative practices, in this view, are explained along the same lines as the use of cultural artifacts:

to socially learn the conventional use of a tool or a symbol, children must come to understand why, towards what outside end, the other person is using the tool or symbol; that is to say, they must come to understand the intentional significance of the tool use or symbolic practice – what it is 'for,' what 'we,' the users of this tool or symbol, do with it. (Tomasello 1999, p. 6).

Experience with perspectival representations mediates each new generation's enculturation into the collective conventions and institutions of a society - that is, the *collective intentionality* of a cul-

THE WORLD WE POINT AT

ture (Tomasello & Rakoczy 2007; Searle 1995). Collective intentionality enables human cognitive resources to be pooled and elaborated on in each new generation, allowing for rapid *cumulative cultural evolution* using the ‘ratcheting effect’ of systematically building on the achievements of previous generations, and which works on time scales many orders of magnitude faster than biological evolution.

A Phylogenetic Account

An explicit evolutionary account of the uniquely human socio-cognitive skills of mind-reading, symbolic communication, the use of artifacts, collective intentionality, and rapid cultural evolution has been formulated in the *Vygotskian Intelligence (VI) Hypothesis* (Moll & Tomasello 2007). On this theory, while primary selective pressure for the evolution of primate cognition in general has been social competition as emphasized by the social intelligence or Machiavellian Intelligence hypotheses (Humphrey 1976; Byrne & Whiten 1988), the evolutionary driving force for uniquely human cognition has been social *cooperation* (see also, Tomasello 2009). Humans are *biologically adapted* for social interactions involving joint attention and *shared intentionality*, encompassing “joint goals, joint intentions and joint attention, and special motivations for helping and sharing with others” (p. 645). Chimpanzee subjects in the object-choice task do not take pointing or gazing as an informative cue for food because primate cognition is not naturally adapted to read cooperative communicative cues. To test this hypothesis, Hare and Tomasello (2004) devised a version of the object-choice task in which a more ecologically valid *competitive* food-access relationship was established between a human and a chimpanzee. After the food was hidden in one of two containers, instead of pointing to the relevant container, the human extended her arm effortfully towards the container where the food was hidden, but not reaching it. The action closely resembled pointing except for the fact that all fingers were extended. In this competitive variation of the object-choice task, chimpanzees spontaneously performed better in choosing the container with the hidden food. In this version of the object-choice task, according to the *VI Hypothesis*, chimpanzees perceive the goal-directedness of the reaching action and infer from this that there is something desirable in the container the action is directed towards. There is independent evidence for chimpanzee skill in this type of exploitation of social cues (Call 2004). But understanding instrumental intentions is not the same as understanding communicative intentions which, they argue, underlies successful understanding of the pointing gesture in the task.

THE WORLD WE POINT AT

According to the *VI Hypothesis*, the evolution of uniquely human cooperative shared intentionality could have occurred through a two stage process: First, the adaptive advantages of a more cooperative socio-ecology targeted variance in emotional reactivity in a ‘self-domestication’ process allowing for more group activities such as foraging, hunting and feeding, with less competition and aggression. And second, this new adaptive space could have allowed for a more direct selection of individual variance in social-cognitive and motivational skills for more diverse and complex cooperative activities involving shared intentionality and cooperative communication (Hare & Tomasello 2004, Moll & Tomasello 2007, Tomasello & Carpenter 2007). Tomasello develops this scheme in subsequent works (Tomasello, 2008; Tomasello, 2009).

The Shared Intentionality Account In A Broader Context

According to Tomasello and colleagues’ the act of referential pointing reveals a variety of unique and powerful social-cognitive abilities for cooperative communication, shared intentionality and – later -the collective intentionality of culture. Their shared intentionality account offers a fruitful, coherent and integrative account of uniquely human cognition, communication & culture, incorporating methodologies and insights from comparative psychology, cognitive psychology, psycholinguistics, developmental psychology, cultural psychology and philosophy. It has contributed to the pragmatic approach to human communication in which the philosopher of language Grice (1957, 1975) laid important foundations with his ideas of cooperative motives for communication, speech acts and the role of communicative intentions in inferring meaning. A large and growing body of research supports the claim that children’s word learning rests fundamentally on their social-pragmatic skills, in which an understanding of the declarative pointing gesture plays an important role (Baldwin & Moses 2001; Saylor, Baldwin & Sabbagh 2004). Recent interest in the *ostensive* nature of communication and knowledge transmission (Csibra & Gergely 2009; Sperber & Wilson 1986) also draws heavily from pragmatic traditions. The shared intentionality account is also resonant with broader interest in role of cooperation in evolution (Nowak 2006), and has clear relevance to the growing interest in social cognition (Carpendale & Lewis 2006), and social foundations of cognition in joint action, shared cognition, and imitation – reflected in the burgeoning interest in mirror neurons, joint action and ‘common coding’ (reviews, Cattaneo & Rizzolatti 2009; van der Wel, Sebanz & Knoblich in press), and in philosophical and psychological interest in ‘we’ intentionality, in which participants have a shared goal and

THE WORLD WE POINT AT

coordinated action roles for pursuing that shared goal (Bratman, 1992; Plotkin, 2003; Searle 1995; Tuomela 1995).

SECTION III

The purpose of this section is to draw attention to four interrelated shortcomings of Tomasello and colleagues' account of pointing that ramify through to the shared intentionality account more generally. My thesis is that each of these problems reflects the fact that this account gives an incomplete treatment of key dimensions of the *intentionality* of human cognition, communication and knowledge.

Three Types Of Intentionality

In its long history, the term intentionality has acquired multiple meanings in different theoretical traditions from the Middle Ages to Brentano and Husserl, modern linguistics, and the philosophy of mind (Jacob 2010; Haye 2008). Common to all is a 'directedness' or 'aboutness' relationship and there are clear groupings. For our purposes we can distinguish between three types of intentionality that commonly figure in cognitive science:

1. *Intentionality (M)*: Intentionality as *motivated behavior*. Here there is a motivational directedness between an *agent* and *goal* as exhibited in motivated behavior.
2. *Intentionality (R)*: Intentionality as *representation* or *referential content*. Within cognitive science, two subtypes of representations have been commonly distinguished (and they are not understood to be exhaustive): (i) Mental states (whether perceptions or thoughts) or gestural or linguistic forms can make reference to *objects, properties, events or states* of affairs in a conceptual way (Clapin 2002; Searle 1983). I will call this subclass *predicational intentionality (Rp)*. Predication involves *affirming* or *asserting* or *judging* something (the 'predicate') as being of or about something else (a 'topic' or 'argument'). A successful act of perception, on this conception, always involves obtaining *conceptual knowledge* of a predicative sort (at the least) – as in '*that* is yellow', or '*that* is a *ball*'. More complex perceptions can be built up from this kind of predication (e.g. the judgment '*the ball* is next to the *bat*'). This conception of the *intentionality (Rp)* of perception is in keeping with traditional cognitive approaches to visual perception (Gregory, 1997; Kahneman, Triesman, and Gibbs 1992). Intentional (Rp) representations may also be (although not necessarily) *propositional*, where propositions are

THE WORLD WE POINT AT

understood as “contents of perceptual and cognitive states that represent the world as being certain ways, and so have truth conditions” (Soames in press, p. 2; see also Soames 2010). (ii) Representational content can also be *ideomotor* based, in which abstract action plans represent *motor-actions* that unfold spatiotemporally - another well established type of representation in the field of cognitive psychology (review, Smith & Kosslyn 2007). I will call this subclass *motor-action intentionally (Rma)*. Note that *intentional (M)* behaviors could be mediated by *intentionality (R)* motor plans, but they need not, as in basic instrumental conditioning. The motor planning that supports constructing a bed from an assembly kit would be an example of *intentionality (Rma)*.

3. *Intentionality (CI)*: Intentionality as *communicative intention* in the sense of ‘I intend for you to know that I intend for you x’ (Moll & Tomasello 2007, p. 6; see also Grice 1957). In the original Gricean formulation that this is based on, what is conveyed in a communicative act (whether a gesture or speech act) is not merely the content of the gesture/utterance (which was generally understood in the *intentional (Rp)* sense) but also the intention of the speaker to convey that meaning, and the intention of the speaker to convey that meaning *by means of* that specific gesture/utterance.

All of three types of intentionality are found explicitly in Tomasello and his colleagues’ writings, although they do not all play a central explanatory role.

Problem 1: Informative Pointing vs Deictic, Referential Pointing

According to the shared intentionality account, apes cannot point because they cannot understand cooperative communicative intentions within a joint attention frame or common ground. The common ground is a ‘joint project’ such as the hiding finding game in the object-choice task. Communicative intentions are defined as “intentions that are not directed at things or behaviors but at another individual’s intentional states” (Moll & Tomasello 2007, p. 6). ‘Intentional states’ in this sense are assumed to include an understanding of perceptions which are (*intentional (R)*) states, but what is important in the communicative intention is not to simply inform or change knowledge states, but to promote some *goal* – i.e. *intentionality (M)*.

THE WORLD WE POINT AT

One problem with this account of pointing is the following: the standard contexts in which declarative pointing is elicited – simply directing and engage joint attention and interest in an object - does not have this ‘helping’ structure, and is less obviously explained by this account. One important context in which declarative pointing occurs is word learning and the acquisition of conventional, symbolic language more generally (Gliga & Csibra 2009). The creation of ‘referential triangles’ by pointing, incorporating a referent the child and adult in a joint attentional interaction, is widely believed to be foundational for human speech acquisition (Butterworth 2003; Leavens, Hopkins & Bard 2005).

Declarative pointing could be argued to be used ‘informatively’ by adults to help infants learn, on the assumption that the adult and child have some joint project of ‘teaching-learning’ together. In general, learning new words depends on contextualized pragmatic information about what the adult is referring to with the new word and pointing is a particularly effective source of such information (Tomasello 2001). But this does not explain the *infant’s* production of declarative in language acquisition contexts – the *infant* is not helping the parent by providing useful information. The vocabulary of a toddler is predicted by responsiveness to, and *production of*, pointing (Brooks & Meltzoff 2008; Volterra et al. 2005).

The evidence suggests that declarative pointing has a distinctive reference-designation (*intentional (R)* function that is not captured by Tomasello and colleagues ‘informative declaratives’ vs ‘expressive declaratives’ taxonomy. Infants typically begin pointing before they begin using language (Carpenter, Nagell & Tomasello 1998). Pre-linguistic Infants have been demonstrated to have *referential* expectations induced by ostensive signals such as direct eye gaze or pointing. An *object* is expected to be found (review, Csibra & Gergely 2009). And if an adult looks and points behind a barrier while also uttering an object name within a 13 month old infant’s vocabulary, the infant expects not just to find an object there, but an object of the kind named (Senju, Csibra & Johnson 2008). When pointing and a word are used together, infants expect them to refer to the same thing. Adult pointing and lexical information work redundantly for children to establish reference, by as young as 13 months (Gliga & Csibra 2009). Pointing is a very powerful cue to referential intent, sometimes masking toddlers’ established word knowledge. In paradigms which set deictic-declarative and lexical cues to reference against each other, pointing dominates at 2 and even 4 years of age (Grassman & Tomasello 2010). Thus declarative pointing has an important ‘we are *intentionally (R)* sharing in identifying and acknowledging a referent together’ function that is not reducible to an ‘I have an *intentional (CA)* state directed at your

THE WORLD WE POINT AT

intentional (M) state in a joint project' function. This emphasis on the *referent* is found in the *deictic* taxonomy for gestures (Section I), where the contrast is between 'deictic' and 'representational' gestures – in which gestures are used for acting things out in a symbolic way.

Problem 2. Neglecting The Importance of Knowledge Acquisition

There is a growing body of evidence that human communication is uniquely adapted to allow the rapid, single episode transmission of *generic knowledge* between individuals through ostensive *intentional (CI)* communication (review, Csibra & Gergely 2009). Csibra and Gergely call the communication system 'natural pedagogy', and build a strong case that human infants are adapted to benefit from it to acquire contentful (*intentional (R)*) cultural knowledge. They hypothesize that this generic knowledge is communicated both linguistically, transmitting propositional knowledge relating general kinds and properties (e.g. 'Aeroplanes fly'), and through direct demonstrations of planned manual techniques that can be learned imitatively. Children learn from adults by unguided observation, but whenever they are the recipients of ostensive communicative acts their mode of learning changes fundamentally. For example, apparently causally useless actions are imitated only when the demonstrator communicates that they are relevant (Gergely, Bekkering, & Király 2002; Brugger et al., 2007). Communicative intentions have been shown to be critical in what is encoded during human infant imitation learning by Southgate and colleagues (2009). When an action is performed in a particular style to toddlers with a clear end state (e.g. mouse hopping to a house), 12 and 18 month year olds tend to imitate only the end state. However, when the information about the end state is communicated to 18 month olds prior to the demonstration, ('the mouse lives in the house'), they imitate the action sequence more – i.e. the *new* information to be learned, not the redundant information that is part of the common ground in the sense of shared *knowledge*.

Moreover, infants have an interpretation bias to preferentially encode the content of ostensive-referential communication as representing generalizable knowledge. In one revealing study (Yoon, Johnson & Csibra 2008), infants of 9 months were more likely to detect a change in an object's *location* than its object specific features (configuration & color) when they observed a non-communicative reaching action towards it, while they showed the reverse pattern if the object was first seen as the target of an ostensive pointing gesture involving eye contact and joint attentional engagement. In the latter case they apparently completely neglected the location of the object despite the fact that the ostensive

THE WORLD WE POINT AT

pointing cue identified the object for them by specifying a location. In this study the pointing gesture facilitates the encoding of visual features that are relevant for general knowledge acquisition – for recognition and generalization – at the expense of transient location information. It has also been shown that *joint attention* (alternate-gaze) cues helps enable 9-month-olds to encode basic object-feature information about a novel object, relative to conditions in which only the object is presented in the absence of joint attention (Striano et al. 2006).

Thus declarative pointing serves a critical generic-knowledge acquisition function that is distinct from the function to provide useful information for some *intentional (M)* goal in a joint action context.

Problem 3. The Meaning of the Referent in Declarative Pointing.

On the shared intentionality account, while pointing to communicate always involves as one component “the pointer inviting the recipient to attend to some referent...the main function of the indicating finger” (Tomasello, Carpenter & Liszkowski 2007, p. 712), communicative intentions within a joint attention frame or common ground are what is most critical in determining the meaning of the referent.

“Pointing always underdetermines the intended referent without some... shared context within which the pointing occurs. Am I pointing at the whole bicycle? Or the special kind of polyvinyl seat? Or the color? Or the metal material it is made of? The possibilities are limitless, and demonstrate... that the pointing gesture can actually indicate radically different perspectives on one and the same perceptual situation.” (Tomasello, Carpenter & Liszkowski, 2007, p. 706).

Apes are unable to understand pointing, on this account because they cannot understand the communicative intentions to provide useful information cooperatively. “A successful act of reference occurs when one individual intends for another to attend to something within some larger communicative context, and the other recognizes this intention and complies with it” (Tomasello, Carpenter & Liszkowski, 2007 p. 711).

But there is a problem with this relevance-fixing account of referential meaning. The Yoon and colleagues (2008) pointing vs reaching experiment (see above) has important implications for *object cognition* in infants in comparison to apes, a critical issue when considering *intentional (R)* signi-

THE WORLD WE POINT AT

ficance of ostensive-referential cues such as pointing. The *meaning* or *intentionality* (*R*) of referential gestures cannot be taken for granted. Infants follow deictic gestures such as gaze shifts or pointing from 3–6 months or earlier (D'Entremont, Hains & Muir, 1997), but up to 9 months, the infant attends to any salient object which stands in his line of sight. Researchers have proposed that gaze or point following is initially an automatic or reinforced response that reflects no referential understanding (Triesch et al. 2006). Gliga & Gergely (2009) have demonstrated, however, that infants of 13 months understand co-reference between words and pointing, providing strong evidence that infants by this age understand their referential nature. But what does this understanding consist in? On the shared intentionality account, it consists in *picking out* a relevant object or property or event. But there is evidence that perceiving the world in terms of objects, properties and events cannot be assumed.

Referential understanding requires at the minimum that an object or object is individuated and classified. Three types of information may be used for object individuation and object identity: spatiotemporal cues (including position, size and volume), object specific feature information (e.g. colors or spatial configurations), and object kind information (e.g. 'animal'). Studies in object perception suggest that a major developmental change occurs between around 10 - 12 months of age (reviews Johnson, Mareschal & Csibra 2001; Xu 1999): Younger infants' default object encoding mode for object individuation and enumeration is spatiotemporal. It is not until around 12-14 months of age that infants consistently use object property or object kind information in conjunction with spatiotemporal cues to do so. Results therefore indicate that infants can independently process object-feature information and spatiotemporal information, and may be unable to process both together until 12 months of age. The Yoon and colleagues study (2008) showed that ostensive communication (*intentionality* (*CI*)) exhibited in declarative pointing could override a default spatiotemporal mode of object encoding in 9 month olds, biasing them to remember object-feature information rather than position information – thereby altering the *intentionality* (*R*) properties of the perceptions involved in attending to the referent.

These result suggests that in the object choice experiments apes, unlike human toddlers, may only be capable of encoding object *location* or *size* information (relevant to object interaction) given the ostensive reaching cue, not anything about object *identity* information. In support of this interpretation, a number of studies indicate that apes default mode of object encoding when they reach and search for objects is spatiotemporal, not object-property or object-kind based (Yoon, Johnson & Csibra, 2009). For incomplete goal-directed actions (e.g. reaching) which culminate in an action on the

THE WORLD WE POINT AT

object (e.g. eating), it is plausible that apes and younger infants simply encode the location (and possibly size for grasping) of the goal object rather than the identity of the object itself. This encoding may function to predict what will happen next, a function that evidence suggests may be subserved by the mirror neuron system – discussed below (Fogassi et al. 2005). For a *communicative-referential* gesture, on the other hand, the identity of the object-referent is more relevant than its location, allowing for future recognition, like-kind generalization, and word learning (Csibra & Gergely, 2009).

There is thus an additional problem for the shared intentionality account. Considering the object-choice task from a *reference* angle, it is unlikely that the ape fails in the task simply because it cannot pick out a *relevant* referent on account of not understanding communicative intentions – saying in effect “‘A bucket. So what? Now where’s the food?’” Tomasello & Carpenter 2007 p. 122). It is likely that the ape does not individuate an object under an object kind category at all. Rather, it processes position/size or action-affordance information, stripped of predicative-conceptual information which is processed elsewhere. In which case, the ‘disambiguating the referent’ account of the function of communicative intentions needs to be modified. Apes may simply not *have* a perceptual referent in the sense assumed in the *Vygotskian Intelligence Hypothesis* – whether an object, an object feature or an event. This is an important *intentionality (R)* issue that needs to be addressed in any account of why apes fail to understand ostensive referential pointing.

Problem 4. Understanding Intentionality (M): ‘Thick’ or ‘Thin’

Problem 3 has obvious implications for the question of whether apes understand others’ perceptions in their social cognition. In his earlier work (Tomasello 1999), Tomasello had hypothesized that only human beings understand each other as ‘intentional agents’, with *internal goals* and *perceptions* and how they relate to each other in intentional action. This kind of mind-reading was thought to be what accounted for many uniquely human social cognitive skills. In the light of experiments conducted by Hare and colleagues reviewed below this hypothesis was rejected, the claim now being that some apes do in fact understand intentions and perceptions (Call & Tomasello 2008; Tomasello, Call & Hare 2003) and what distinguishes human from ape cognition is in fact *shared intentionality* - the ability “to create with others in collaborative interactions joint intentions and joint attention” (Tomasello, 2006, p. 517). By implication, what prevents apes from understanding the meaning of declarative pointing gestures is not an inability to understand that a conspecific can *perceive* a referent (*intention-*

THE WORLD WE POINT AT

nality (Rp)– but an inability to jointly attention a referent in the context of a common ground that is shared.

The experiments of Hare and colleagues (Hare et al. 2000; Hare, Call & Tomasello 2001) motivated this change of position made use of a ‘food competition’ paradigm in which a subordinate (S) and a dominant (D) chimpanzee were placed in competition over food items that were placed either in full view of both of them, or sometimes behind a physical barrier such that S could see the food but not D. In this setting S flexibly pursued the food that was concealed from D, suggesting S knows what D can and cannot see. In another set of experiments, there were two barriers and the food was placed behind one of them, out of view of D. In the control condition both D and S watched the food hiding process while in the experimental condition only S watched. S pursued the food more often in the experimental condition, suggesting that S knows what D has or has not seen in the immediate past. Moreover, if D is switched for another D (2) after watching the hiding process, S pursues the food more than when there is no switch, suggesting S knows who (D1 or D2) knows what.

The control conditions were devised to counter the possibility that subordinates based their food choices on behavior-reading rules such as <do not feed in the open when a dominant male is present>, or <do not choose the food that the dominant male has oriented or looked towards>. Subsequent experiments have elaborated on this paradigm, in one case with an additional test for false belief understanding (Kaminski, Call & Tomasello, 2008). Tomasello and colleagues conclude from this body of research that although apes do not understand how false beliefs can determine behavior, they “know something about the content of what others see and, at least in some situations, how this governs their behavior” (Tomasello, Call & Hare, 2003, p. 155).

The hypothesis that apes can understand that unobservable mental states may determine behavior has consistently been challenged by Povinelli and colleagues who reason as follows: Most people would agree that the chimpanzee’s mind contains mental representations in the sense of ‘intervening variables’. But “the question is: are these intervening variables representations of behavioral abstractions and mental states (as theoretical entities), or behavioral abstractions alone” (Povinelli & Vonk, 2003 p. 158). They argue that in the food competition and subsequent studies, a behavioral invariant is always confounded with the ‘mentalist’ attribution that is proposed. For example, in the location switching control: <if a dominant conspecific orients to a piece of food in a particular *location*, then

THE WORLD WE POINT AT

that food must be avoided>; in the dominant switching control: <X was present and oriented; he will probably go after the food. Y was not present; she probably won't>.

There is also a plausible alternative to 'rich' mental state attributions and 'lean' behavior abstraction interpretations, found in the mirror neuron literature. Mirror neurons discharge not only when the monkey executes a goal-related action like grasping an object but when observing *other* individuals executing similar actions (review, Rizzolatti & Craighero 2004). Gallese and Maria Umiltà (2006) argue that mirror neurons mediate a direct form of action understanding by means of an automatic *embodied simulation* of observed actions in the observer's own motor system. They function not only to recognize actions but also in 'understanding' – in an automatic, non-conceptual sense - the *intention* promoting the action. This is achieved by the activation of statistically related chains of motor schemata culminating in the distal goal state action (e.g. putting food in mouth), thus coding not only an observed motor act (e.g. grasping food) but those distal motor acts that would normally follow (e.g. putting food in mouth). Such a system can mediate the ascription of *goals* and an anticipation of what the agent is going to do next. Understanding intentions and goal directed behavior as embodied simulation in this way does not require "explicit thinking about the contents of someone else's mind" (p. 28), but nonetheless cannot be reduced to simple behavior-reading of patterns of sensory input.

On this mirror neuron account, there may be understanding of the *intentionality (M)* of behavior, without an understanding of *intentionality (R)*. In order to trigger monkey mirror neurons, unlike human mirror neurons, the executed or observed action must be *transitive* or object directed: both the effector and object must *interact*. Assuming mirror neurons are the neural substrate mediating the understanding and prediction of goal directed actions among conspecifics, and assuming the ape mirror neuron system is like the monkey's, this *transitivity* property implies that an object (such as item of food) may not be represented non-human primate cognition independently from the action *on* it – as an 'object-as-acted-on' *affordance*. Thus rather than inferring from the food competition data that chimpanzees "know something about the content of what others see", a plausible explanation can be constructed in which the subordinate (S) chimpanzee's food choices are the result of an affordance-detecting mirror neuron, interacting with gaze detection attentional mechanisms ((Shepherd et al. 2007), and working memory (Iacoboni et al. 1999)). Under an embodied simulation interpretation, in no conditions of these experiments is it clear that an understanding of psychological contents, the *intentionality (R)*, of mental states, is required. Moreover this 'embodied simulation' account is consistent

THE WORLD WE POINT AT

with, studies directly looking at object individuation and identity which call into question whether apes perceive spatiotemporal and object-kind information as an integrated whole. The default spatiotemporal mode of ape object individuation could be supported by an affordance-encoding mirror neuron system.

The food competition findings as they stand, in conclusion, do not warrant confidence in the inference that apes' understand the *intentionality (R)* of perception - that what other conspecifics or humans *perceive* and *know about* are objects, events or states of affairs in the world. Therefore, it does not warrant explaining ostensive referential pointing in terms of determining 'referential relevance' via communicative intentions and common ground.

Summary

The four problems identified above have a common thread: that the *referential* nature of gestures and other ostensive communicative cues, and the *representational* nature of mental states, is problematic for Tomasello and colleagues' account of pointing, and for the shared intentionality account more generally. For this reason, a social-pragmatic account of uniquely human cognition, as revealed in declarative pointing, must do more than explain successful ostensive acts of reference in terms of constraining possibilities via communicative intentions and common ground. It must explain how the *intentionality (R)* arises in the first place – how objects and their properties and events can be perceived *as* objects, properties, and events, rather than action-affordances or dissociated spatiotemporal and conceptual object-kind information. In loose terms, the *semantics* in uniquely human cognition is as important as the *social pragmatics*.

SECTION IV

The Shared Intentionality (R) Hypothesis

Any adequate account of the ontogeny and phylogeny of communicative pointing, and what it reveals about uniquely human cognition, needs to incorporate (at least) three types of intentionality - *intentionality (M)*, *intentionality (R)* and *intentionality (CI)* - within a single explanatory framework. The s joint action framework of Tomasello and colleagues does not, I have argued, adequately account for *intentionality (R)*. In this section I develop a theory that builds on the central role ascribed to

THE WORLD WE POINT AT

shared *intentionality (M)* and *intentionality (CI)* by Tomasello and colleagues, but which also gives a central role to *intentionality (R)*.

Csibra and Gergely hypothesize that the emergence of recursive tool making practices “confronted the observational learner with cognitively opaque contents to acquire” (Csibra & Gergely 2009, p. 148) that could not be transmitted from generation to generation using existing observational learning mechanisms. A learning system based on ostensive-referential demonstrations of knowledge evolved for the transmission of this kind of structured ‘opaque’ knowledge, ultimately accounting for both generic skill knowledge and generic conceptual knowledge - i.e. ‘generalizable kind-relevant information’ (Csibra & Gergely 2006). The *Shared Intentionality (R) Hypothesis* builds on this premise, classifying the representations supporting human imitation learning and joint action as *intentional (Rma)*, and those supporting perception, propositional thought, and the transmission of generic cultural knowledge, as *intentional (Rp)*. The hypothesis holds that both of these types of representation (along with the thinking associated with them) are uniquely human, and function together synergistically in joint action, imitation learning, pedagogy and language. They are both reflected in infant acts of declarative pointing, and they depend on the co-evolution of:

1. A uniquely human functional integration of the dorsal ‘action’ and ventral ‘perception’ visual pathways allowing for indexical, situated object and event perception.
2. A socio-cognitive *joint attention* system that evolved to facilitate pragmatic ostensive-communication (*intentionality (CI)*) of (shared) *intentional (R)* representations in cultural cognition.
3. A uniquely human mirror neuron system, integrating communicative and action-coding primate mirror functionality, that supports both common-coded (shared) *intentional (Rma)* and *intentional (Rp)* representations.

The *Shared Intentionality (R) Hypothesis* is depicted in Figure 1.

THE WORLD WE POINT AT

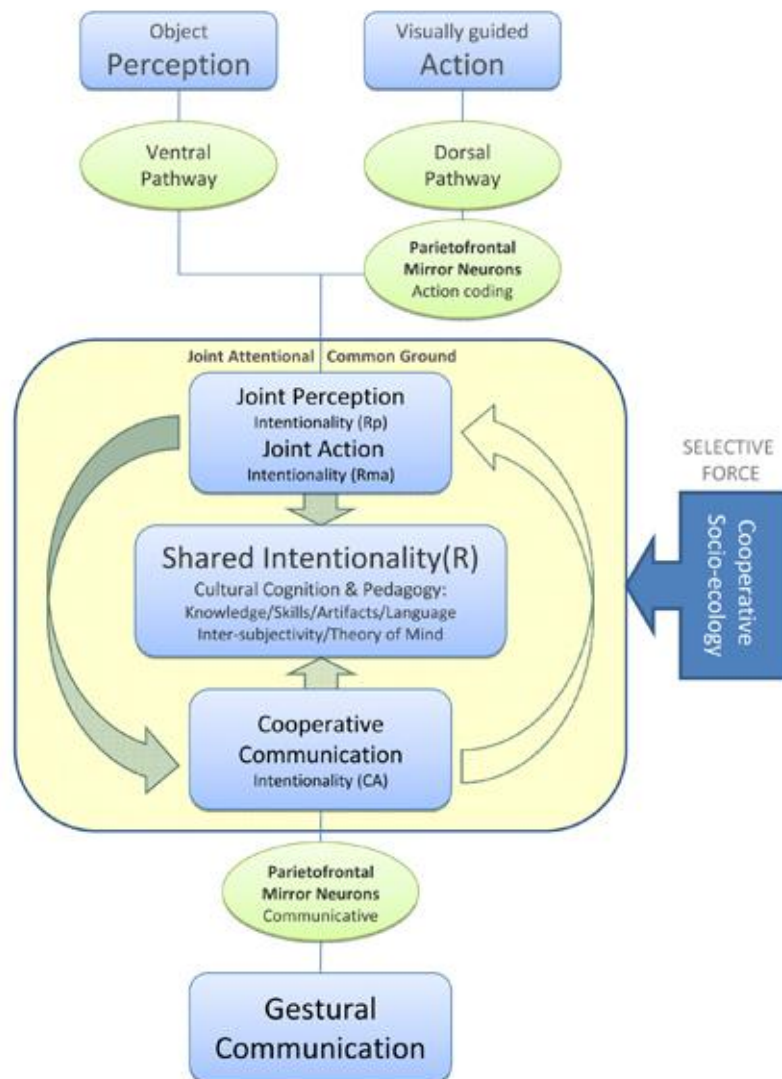


Figure 1. The *Shared Intentionality (R) Hypothesis* for the Evolution of Uniquely Human Cognition

In support of a central role of *intentional (Rma)* representations in uniquely human cognition, human children (and adults) are experts at abstracting away from the particulars of observed behaviors, focusing instead on sub-goals achieved along the way to solving a given task (Gattis, Bekkering & Wohlschlagler 2002). Human do this in both in local coupled interactions and in broader motor-planning contexts such as navigation or coordinating joint projects, and it is critical to event perception more generally (Kurby & Zacks 2007). There is fieldwork evidence for the existence of recursive motor skills in some great apes (Byrne & Russon 1998), and some experimental evidence for goal directed action planning in problem solving (Hihara et al. 2003). But the experimental evidence for spontaneous and flexible multi-stage motor planning as we see in human children is weak. There is evidence for

THE WORLD WE POINT AT

some imitation learning ability in chimpanzees (Whiten, Horner, & de Waal 2005; Buttelmann, et al. 2007), but there is no evidence for rapid imitation learning of multi-step procedures as we find in children (and adults). Evidence for instructed learning of instrumental actions in non-human primates in which a ‘teacher’ adjusts his behavior contingent on the learner's progress in skill or knowledge is also rare or non-existent (Boesch & Tomasello 1998).

As for the central role of *intentional (Rp)* representations, the power and ubiquity of human spoken language is the most striking evidence: language is present in virtually all humans, is supported by dedicated neural circuitry, has a characteristic development pattern, and is grounded in a number of constraints that can be characterized by universal linguistic forms (such as ‘noun phrase’) (Hauser & Fitch, 2003). None of this is true of other species. In addition, cognition of an inferentially productive, propositional nature, in which objective but unobservable explanatory variables such as *causal forces* and *beliefs* in theory of mind, as well as abstract higher order conceptual relations more generally, is also plausibly uniquely human (Penn, Holyoak & Povinelli 2008).

Any account of cumulative cultural evolution must take account of both the vast cultural store of techniques and skills with an *intentional (Rma)* basis that are learned, exchanged and modified in each generation, as well as the vast store of general knowledge of a propositional, factual *intentional (Rp)* type.

Dorsal And Ventral Pathway Integration

In Section III I argued at length that the representational/referential aspect of the shared intentionality account of declarative pointing was inadequately explained. This section now provides such an explanation. According to the *SI(R) Hypothesis*, referents for acts of pointing, as well as linguistic symbols or structures, can fall into two broad classes: (i) objects (nouns) and their properties or relations, and (ii) actions (verbs) between objects or agents/patients. Both have a predicative (argument-assertion) structure, and both depend on a functional integration of two well-recognized visual pathways. Extensive anatomical, physiological, neuropsychological and behavioral evidence (review, Milner & Goodale 2008) supports a structural and functional dissociation of the primate cerebral cortex between a dorsal ‘action/where’ pathway and a ventral ‘perception/what’ pathway:

The ventral stream transforms visual inputs into perceptual representations that embody the enduring characteristics of objects and their spatial relations. These representations enable us to parse the scene, and to think about objects and events in

THE WORLD WE POINT AT

the visual world. In contrast, the dorsal stream's job is to mediate the visual control of skilled actions, such as reaching and grasping, directed at objects in the world (p. 774).

The ventral pathway projects from the retina through to inferior temporal cortex, with further projections to ventral prefrontal cortex. The dorsal pathway extends from retina through to the posterior parietal cortex to the dorsal premotor cortex (Van Essen, Anderson & Felleman 1992; Rao, Rainer, & Miller 1997).

Modifying a hypothesis proposed by Hurford (2003), with similarities to a line developed by Jackendoff and Landau (1992), the *SI(R) Hypothesis* claims the following. Every logical scheme has at its heart an asymmetry between two types of term of the form PREDICATE (X) which differ in their *semantics* (how they relate to the world), and in their *syntax* (how they relate to each other in the formal scheme). The arguments denote individual entities, while predicates denote classes or properties. The PREDICATE (X) formula as a whole represents a state of affairs or event in the world. The claim is that *intentional (Rp)* perceptual representations instantiate this syntactic structure, in which the argument is the spatiotemporally located object that is the focus of attention (*that* which I am oriented towards and may act on), and the predicate is some property or conceptual category attributed to it (e.g. redness or the general category 'ball'). The integration of the two allows for what Pylyshyn calls "a direct, preconceptual connection between objects in the visual world (visual objects or proto-objects) and their representations in the visual system" in his 'situated cognition' account of vision (2000, p. 197). An obvious mapping in visual perception of objects presents itself: the former is encoded in the dorsal 'action/where' pathway; the latter in the ventral 'perception/what' pathway. These *intentional (Rp)* representations are proposed to be *proto-linguistic* – designed to be mapped onto linguistic (or representational-gestural) structures. As argued below, they evolved in the context of ostensive communication, and the neural integration was selected for to facilitate *communication*. Mechanisms for the further elaboration of the basic PREDICATE (X) structure into more complex and inferentially productive *linguistic* representations are discussed by Hurford (2003) and have been explored by formal semanticists (e.g. Parsons 1990). While there is much more to human language than predicate-argument structure, this perceptually based structure is taken to be the semantic foundation on which the rest is built. Arguments are typically instantiated in pronouns or proper names, while predicates map to common nouns, verbs, adjectives, adverbs and prepositions.

THE WORLD WE POINT AT

Action-based predicative constructions (such as ‘she threw the ball’) underlying *event perception* (Kurby & Zacks, 2007) can be interpreted in terms of action predicates being ascribed to object arguments. Aside from ‘anchoring’ a particular object via a deictic indexing to in order to predicate a conceptual attribute, the dorsal pathway can also support ‘event structure templates’, that map linguistically onto verb constructions (Kemmerer et al. 2008). According to the *SI(R) Hypothesis* this involves the recruitment of motor action plans that evolved in interaction with object-based *intentional (Rp)* perceptual representations (their ‘arguments’), and to a large extent are suited for mapping onto *intentional (Rp)* linguistic structures. This has strong empirical support in brain imaging studies showing dorsal stream - particularly Broca’s area – activation while making semantic judgments about verb meanings (Kemmerer et al. 2008).

What is central to this PREDICATE (X) hypothesis is the role of declarative pointing as a deictic, ‘argument’ (X) identifying gesture – with the referent being the located entity (and a possible site of action) and its object category or set of features being the (conceptual) predicate/s. (On this interpretation, the ‘sharing of affect or attitude’ component in traditional proto-declarative pointing could be interpreted as a pre-cursor of a ‘propositional attitude’ - the person’s attitude (such as affirmation or doubt or desire) towards a proposition.) Declarative pointing, on this account, reflects a fundamental – and uniquely human – ability to individuate an object in space and time, and classify it predicatively. Conditional on this, action-events involving objects may be declaratively pointed to as well.

There is evidence for a central role of predicate-argument structures in early language development involving pointing. When human infants first begin talking, many of their earliest utterances are combinations of gestures (mostly pointing) with words, which divide up in various ways the topic and focus functions (Tomasello 2003a). The first acts of declarative pointing involving naming can be construed as a predicative and initially in development only the predicate - a word (noun, adjective, verb) - is an obligatory element in a such *intentionality (Rp)* constructs: the referring (as opposed to the ‘commenting’) part of the construct is implied by ostensive-communicative signals such as pointing (Allen, 1973). According to Iverson and Goldin-Meadow (2005), toddlers subsequently start using ‘supplementary combinations’ by e.g. pointing to a dog and saying ‘sleep’ to convey the *intentional (Rp)* message that ‘the dog is sleeping’ about 4 months before they are able to combine two words to express that same meaning (e.g. ‘dog sleeps’).

THE WORLD WE POINT AT

Evidence for a close link between *intentionality (Rp)* and *intentionality (Rma)* in development includes the following. To produce syntactic constructions, children must be able to produce a multiple step procedure towards a single goal, assembled conceptually ahead of time. This is possible non-linguistically in motor planning from 14-18 months, at the same time imitation learning of sequential actions of others becomes possible (Bauer, Hertsgaard, & Dow 1994). Recent handedness studies (Vauclair & Imbault 2009) are also revealing. There is a clear right-handed (and left hemisphere) preference from infancy for both manipulative actions and pointing gestures. While there is no within-individual correlation of hand preference between communicative gestures and bimanual object manipulations in chimpanzees and baboons, for humans, there is a very strong correlation – but only at two key phases of language acquisition: the ‘vocabulary spurt’ at 18-20 months, and the growth spurt for syntax at 29-32 months (Vauclair & Imbault 2009). This suggests that while a left lateralized gestural-communication system within the ape brain is not functionally integrated with the brain system used in object manipulation, these systems are functionally integrated in humans, at least during critical periods of linguistic development where there would be expected to be high cognitive load in working memory. Moreover, given the critical role of joint attentional communicative signals during these periods, *joint attention* brain circuitry would be expected to play a central role in this integrative process.

Joint Attention and Communicative Intentions

According to the *SI (R) Hypothesis*, the functional integration of dorsal and ventral pathways for PREDICATIVE (X) representations evolved in mutual-reinforcement with uniquely human pre-frontal attention joint-attention system that supported ostensive (*intentional (CI)*) communication. This claim is distinct from Hurford’s who believes that predicative representations are common to all primates and possibly other mammals. Indexicality and reference, on my account, have perceptual, gestural and linguistic versions, sharing the concept of joint attention. An act of declarative pointing on this model is a paradigm case of joint-attention to a deictic object that is predicated of, in a way that ultimately supports generic knowledge transmission in natural pedagogy (Csibra & Gergely 2009) as well as language acquisition. In Section III we reviewed ways in which joint attention

Joint-attention behaviors in infancy fall into two categories: responses to attentional orienting of others such as gaze following, and spontaneous initiations of joint attention where infants use ostensive referential gestures and eye contact to direct others’ attention to objects, events, and themselves

THE WORLD WE POINT AT

(Mundy et al. 2007). These behaviors are subserved by two interacting attention-regulation systems (review, Mundy & Newell 2007): (1) a *posterior* ‘orienting’ and perceptual attention system that automatically orients attention towards biologically meaningful stimuli; (2) a later developing, pre-frontally mediated *anterior* system controlling volitional, goal-directed attention allocation that is constrained by self-appraisal of behavior. Imaging data indicates that the posterior attention system is well represented in primates but that the anterior (pre-frontal) system is uniquely human.

Infants only begin to integrate location and conceptual classificatory information for object individuation at the time that they develop their joint attention skills as reviewed above. By 14 months of age most humans have the capacity to obtain the visual attention of another and to redirect it to distal objects, locations and events of interest by pointing. This is the same time that they can individuate objects that disappear and appear from a screen based on both color and location (Johnson, Mareschal & Csibra 2001). Strong interactions between infant referential pointing in (presumably) joint attentional contexts and dorsal pathway mediated manual actions during key cognitive-linguistic developmental periods also supports this hypothesis. In addition, the anterior projection sites of the ventral and dorsal visual pathways (dorsal and ventral prefrontal cortex) are exactly the same as two of the three cortical sites subserving the ‘initiating joint attention’ (IJA) circuitry identified by Mundy and colleagues (review, Mundy & Newell 2007).

Since intentional (*Rma*) motor action representations evolved in tandem with object-based *intentional* (*Rp*) perceptual representations, they also depend on joint attention mechanisms being in place. Moreover, imitation learning, as well as joint action games and problem solving, unfold in joint attention communicative contexts. Sebanz, Bekkering, & Knoblich (2006), have argued that knowing *what* another individual is attending to in a particular joint action context provides important cues about their action goals. They also point out that imitation learning plays an important role in cultural transmission of the *functional meaning* of objects. Caregivers show infants objects to transfer knowledge *about them* via imitation games, such as shaking a rattle while engaging in joint attention and encouraging the infant to do the same – to learn about the properties of rattles and develop *intentional* (*Rp*) representations such as ‘rattles make a rattling sound’.

THE WORLD WE POINT AT

Action-Encoding And Communicative Mirror Neuron Systems

The enabling role of joint attention in ostensive-communicative contexts such as declarative pointing implies that the *intentional (Rp)* contents of visual perceptions, as well as *intentional (Rma)* representations of skills and planned actions, are inherently *shared* (or *shareable*) as well as *intentional (CI) communicable* within some ‘common ground’. In this section we provide an account of this representational sharing and communicative intentionality.

On the *SI(R)* account, a primate-specific neural system that has substantial shared circuitry with both the dorsal visual pathway and posterior joint attention circuitry, plays a central role in both (i) common coding for of *intentional (Rma)* representations and *intentional (Rp)* representations; and (ii) ostensive communication (*intentionality (CI)*). This is the parietofrontal *mirror neuron* system (reviews, Cattaneo & Rizzolatti 2009; Rizzolatti & Craighero 2004).

As reviewed in Section III, one class of monkey mirror neurons discharge not only when the animal *executes* a goal-related action like grasping an object but also when *observing* other individuals executing similar actions, and mirror neurons have thus been widely claimed to function in anticipating or predicting goal directed action in conspecifics in a form of ‘embodied simulation’ social cognition (Rizzolatti & Craighero 2004).

Another major class of primate mirror neurons has a communicative function (Rizzolatti & Craighero 2004). Recent brain imaging evidence shows that gestural signaling activates ‘Broca’s’ homolog in chimpanzees – a premotor area that is part of the primate mirror neuron system (Tagliapietra et al. 2008). Communicative gestures are only found in apes and humans. They are population specific, flexible with context dependent meanings, under greater cortical control than affectively expressive facial and vocal signals, and have a right hand and left hemisphere bias for both apes and humans (review, Pollick & de Waal 2007). A large body of evidence supports *gestural* origins of human spoken language (review, Fogassi & Ferrari 2007). But ape gestures are used to elicit behaviors and achieve social goals, not to impart generic knowledge. Ape gestures, unlike spoken language, are not symbolic, conventionalized signaling systems. Ape gesturing is clearly a precursor of *ostensive communication* of the sort found in humans (*intentionality (CI)*), nonetheless, in as far as it is intentionally communicative between individuals (Bates, Camaioni & Volterra, 1975) and involves intentionally adjusting the gesturing action to match the attentional states of recipients in an ostensive-communicative way (Povinelli et al., 2003b). Pre-frontal mirror neuron circuitry subserving gesturing in apes thus interacts with

THE WORLD WE POINT AT

known attention (gaze) mirror neurons in the superior temporal sulcus and the parietal cortex (Bristow, Rees, & Frith 2007; Shepherd et al. 2007) - components of *posterior* joint attention neural circuitry supporting the kind of gaze alternation shown by apes when they point imperatively for caretakers (Leavens, Hopkins & Bard, 1996).

The human mirror neuron system, unlike the primate system it evolved from, plays a primary role in imitation learning (Molenberghs, Cunnington & Mattingley 2009), symbolic language (Fogassi & Ferrari 2007), and coordinated joint action (Newman-Norlund et al. 2007). Brain imaging experiments in humans have demonstrated that Broca's area, an important language area, is involved in complex finger movement tasks, and hand imitation tasks; and mirror neurons during imitation learning closely interact with dorso-lateral prefrontal cortex neurons – an area implicated in the anterior joint attention circuit (review, Rizzolatti & Craighero 2004). According to the *SI(R) Hypothesis*, over the course of hominin evolution, there has been increasing functional integration of the anterior-posterior joint attention anterior circuitry with two major classes of non-human primate mirror neurons: those encoding motor actions (Jeannerod 1994; Rizzolatti, Rogassi & Gallese 2001) and those supporting communicative gesturing (Rizzolatti & Arbib 1998). We can speculate that this integration of neural systems accompanied the species-wide emergence of deictic referential gestures such as declarative pointing in early infancy. Given that natural pedagogy, human language, and joint action depend on social *cooperation*, we can assume that the adaptive advantage of group cooperation was central to the evolutionary process, as proposed by Tomasello in his Vygotskian Intelligence Hypothesis (Moll & Tomasello 2007).

What is Shared? Common Coding For Intentionality (Rma) and Intentionality (Rp)

With reference to the issue of what is *shared* in joint attention, this mirror neuron account is consistent with numerous behavioral studies showing that action observation activates a corresponding representation in the observer's motor system – i.e. there is a *common code* for perception and action (review, van der Wel, Sebanz & Knoblich in press). It has been shown that goal directed task performance can be distributed between more than one individual via the common coding platform such that the actions one performs and actions one observes in others are represented in a functionally equivalent way. Actions at another person's command are represented just as if they were at one's own command, even if in laboratory settings they result in a decline in one's own performance (Sebanz, Knoblich, &

THE WORLD WE POINT AT

Prinz 2003). Common perception-action coding has been shown to enable judgments about others' goal directed actions (a mind-reading ability) and efficient coordination of joint action - something ubiquitous in human social life (reviews, Knoblich & Sebanz 2006; van der Wel, Sebanz & Knoblich in press). These ideas are all consistent with the *SI(R) Hypothesis*.

Perception, however, may be 'common-coded' in two distinct ways on the *SI(R)* account: (i) with shared *intentional (Rma)* representations in visually guided *action*, allowing for coordinated joint action (as investigated in the common coding literature); but also (ii) with shared *intentional (Rp)* visual *perceptions* (from different gaze-indexed perspectives) allowing for 'joint perception' in the sense of shared predicative knowledge about objects and their properties. It is known that people respond more quickly to targets that are the object of another's attention, even when this ostensive cue is brief or consistently misleading (Friesen & Kingstone 1998). Recent intriguing studies by Samson and Apperly (2010) support the notion of common coding of *intentional (Rp)* perceptions. They showed that adults implicitly compute what someone else can and cannot see. In a task requiring a perceptual judgment of the number of red disks on a wall, participants' judgments were slower and more error-prone when the agent had a different perspective (providing different knowledge) to theirs (Samson et al. in press). An *intentional (Rp)* 'common coding' account is also consistent with evidence for infants' false belief understanding at 18-24 months in anticipatory looking or active helping paradigms (Buttelmann, Carpenter & Tomasello 2009; Southgate, Senju & Csibra 2007). This framework may also contribute to the debate concerning why classic verbally based tests are passed much later (Apperly & Butterfill 2009). As discussed above, one can envisage how early proto-linguistic *perceptual* predicative representations may be abstracted through the course of development into inferentially productive, linguistic structures (with explicit belief attributions of a truth-functional propositional sort), and how in the early stages of this linguistic development a strong egocentric bias prevails. This 'common coding' hypothesis for *intentional (Rp)* perception in both infants and adults could be investigated in future using paradigms analogous to those used in traditional common coding action research.

Common coding research has been represented as reflecting an 'embodied cognition' view in contrast to the traditional view of the mind as a 'symbol manipulator', where "perception and action are understood to be grounded in the physical properties of our bodies as well as the external world we interact with" (van der Wel, Sebanz & Knoblich in press). It has also been portrayed as an investigation into 'lower level' (sometimes called 'resonance') processes related to action anticipation and

THE WORLD WE POINT AT

coordination in social cognition, that are dissociable from but which may complement ‘higher level’ thinking about and communicating with others linguistically (Koblich & Sebanz 2006). This conception finds a parallel in widespread ‘embodied simulation’ or ‘direct resonance’ interpretations of mirror neuron function (although for a leaner interpretation see Csibra, 2005). Gallese and Umiltà hypothesize that not only monkeys but also humans use a basic functional mechanism of embodied simulation which provides “direct access to the intentional states of others” (2006 p. 28). “In our social transactions,” they observe, “we seldom engage in explicit interpretative acts. Our understanding of social situations most of the time is immediate, automatic and almost reflex-like” (p. 25). This conception is also consistent with Gomez’ account of referential pointing in young infants and apes. According to Gomez behaviors may be directly perceived as ‘intentional’ in the sense of being ‘directed to things other than themselves’ but without necessarily representing that they are driven by unobservable mental states; just as with pointing, “understanding that gaze is directed to an object does not require attributing the mental experience of seeing the object —such directionality is directly attributed to gaze itself” (Gomez 2007). We can call this fourth type of intentionality – *intentionality (A)*.

It is possible that there are *two* modes of common coding: the *intentional (R)* type for both perception and action proposed by the *SI (R) Hypothesis*, and an *intentional (A)* type that is pre-conceptual and implicit (i.e. that is not *intentional (R)*), possibly shared in common with non-human primates. But according to the *SI (R) Hypothesis* the former evolved from the latter - possibly when mirror neurons developed *intransitive* properties (Rizzolatti & Craighero 2004), enabling representational coding in the absence of goal objects. On the *SI (R)* account, *intentional (A)* common coding may be apparent in early infancy, prior to joint-attentional ostensive communication . This notion is consistent with evidence that default code for individuating objects before around 9 months is spatio-temporal and not conceptual based (Johnson, Mareschal & Csibra 2001), that 6 month old infants are able to infer the size of a goal object from an actor's grasping movement (Daum et al. 2009) as well as the goal of an uncompleted action from an allocentric perspective (Daum, Prinz & Aschersleb, 2008) However, after 12-14 months, when joint attention skills are well established, the *SI (R)* account claims that common-coding is predominately *intentional (R)*, not *intentional (A)* – albeit *perceptually* based, rather than linguistically based at first.

A ‘higher level’ route to automaticity is possible on this account: there is extensive evidence for ‘dual process’ *automatic* and *controlled* processing (Briggs & Johnsen 1973) in both perception

THE WORLD WE POINT AT

and action, with the former being fast and parallel and the latter slow, effortful and attention-demanding. These two processing modes are functionally seamless, with automatic processing depending on controlled ‘higher level’ processing in the service of skill acquisition (Schneider & Shiffrin 1977). This type of learned automaticity would be expected to play a central role in both imitation learning and joint action.

The World We Point At

By way of conclusion, we may quote from Tomasello’s earlier work:

After they understand others as intentional agents like themselves, a whole new world of inter-subjectively shared reality begins to open up. It is a world populated by material and symbolic artifacts and social practices that members of their culture, both past and present, have created for the use of others. (Tomasello 1999, p. 91).

According to the *SI(R) Hypothesis*, the reality that children are enculturated into is not opened up solely or predominantly by virtue of understanding others as *intentional (M)* agents like the self and being able to collaborate and share in that intentionality. It is shared by virtue of the common-coded representational *intentionality (R)* of both action and perception. The uniquely human *shared intentionality (R)* of perception and action *externalizes* in our cognition an ‘objective’ world of particular objects that we can perceive and interact with, that are inherently *communicable*. From a *deictic*, indexical basis in early infancy that is manifest in joint-attentional communicative acts of declarative pointing, this type of perceptually grounded cognition is further abstracted and elaborated into a linguistically mediated, and highly creative cultural world of shared semantic knowledge and skills, techniques and procedures that can flow through the generations. And as a corollary to this *ontological* externalization in human cognition, the shared *intentionality (R)* of action, perception and thought *internalizes* in our cognition an inter-subjective world populated by embodied minds who are both *agents* with goals and intentions and epistemic *subjects* of knowledge and belief (for a related conceptualization, see Eilan, 2005). This intersubjectivity is first manifest in infants’ joint attentional declarative pointing.

The argument developed in this paper is that the representational *intentionality (R)* of cognition is not given the kind of explanatory account it deserves in Tomasello and colleagues’ social-pragmatic shared intentionality account, and this reflects in the identified shortcomings in of their ac-

THE WORLD WE POINT AT

count of infant declarative pointing. The *SI(R) Hypothesis* builds on the obvious strengths of the shared intentionality account of both declarative pointing and uniquely human cognition – including the emphasis it places on communicative intentions, joint attention, referential triangles, perspectival cognition, and shared intentionality considered as cooperatively motivated joint action. But it gives a much greater explanatory role to a traditional preoccupation of cognitive scientists: the representational *intentionality (R)* of cognition.

References

- Allen, D. A. 1973. The development of predication in child language. *Final Report. National Center for Educational Research and Development (DHEW/OE)*, Washington D.C.
- Apperly, I. A., & Butterfill, S. A. 2009. Do humans have two systems to track beliefs and belief-like states? *Psychological Review*. 116, 4 pp. 953-970.
- Baldwin, D. A. and Moses, L. J. 2001. Links between social understanding and early word learning: challenges to current accounts. *Social Development*, 10, 3 pp. 309–329.
- Bard, K. A. 1992. Intentional behavior and intentional communication in young free-ranging orangutans. *Child Development*, 63, 5 pp. 1186-1197.
- Bates, E. 1979. *The Emergence of Symbols: Cognition and Communication in Infancy*. New York: Academic Press.
- Bates, E., Camaioni, L. and Volterra, V. 1975. The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly*, 21 pp. 205-224.
- Bauer, P. J., Hertsgaard, L. A. and Dow, G. A. 1994. After 8 months have passed: long-term recall of events by 1- to 2-year-old children. *Memory*, 2, 4 pp. 353-82.
- Behne, T., Carpenter, M. and Tomasello, M. 2005. One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science*, 8 pp. 492-499.
- Bernardis, P., Bello, A., Pettenati, P., Stefanini, S. and Gentilucci, M. 2008. Manual actions affect vocalizations of infants. *Experimental Brain Research*, 184 pp. 599-603.
- Boesch, C. & Tomasello, M. (1998). Chimpanzee and Human Cultures. *Current Anthropology*. 39, 5 pp. 591-614.
- Bratman, M. E. 1992. Shared cooperative activity. *Philosophical Review*, 101, 2 pp. 327–341.
- Briggs, G. E., and Johnsen, A. M. 1973. On the nature of central processing in choice reactions. *Memory and Cognition*, 1 pp. 91–100.
- Bristow, D., Rees, G. and Frith, C. D. 2007. Social interaction modifies neural response to gaze shifts. *Social Cognitive and Affective Neuroscience*, 2, 1 pp. 52-61.

THE WORLD WE POINT AT

- Brooks, R. and Meltzoff, A. N. 2008. Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, 35 pp. 207–220.
- Brugger, A., Lariviere, L. A., Mumme, D. L. and Bushnell, E. W. 2007. Doing the right thing: infants' selection of actions to imitate from observed event sequences. *Child Dev.* 78, 806–824.
- Bruner, J.S. 1975. The ontogenesis of speech acts. *Journal of Child Language*, 2 pp. 1-9.
- Buttelmann, D., Carpenter, M., Call, J. and Tomasello, M. 2007. Enculturated chimpanzees imitate rationally. *Developmental Science*, 10, 4 pp. 31-38.
- Buttelmann, D., Carpenter, M. and Tomasello, M. 2009. Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, 112 pp. 337–342.
- Butterworth, G. 2003. Pointing is the royal road to language for babies. Kita, S. (ed.), *Pointing: Where Language, Culture, and Cognition Meet* (pp. 9–33). Mahwah, NJ: Erlbaum.
- Byrne, R. W. and Russon, A. E. 1998. Learning by imitation: A hierarchical approach. *Behavioral and Brain Sciences*, 21 pp 667-684
- Byrne, R.W. and Whiten, A. 1988. *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans*. Oxford: Oxford University Press.
- Call, J. 2004. Inferences about the location of food in the great apes (*Pan paniscus*, *Pan troglodytes*, *Gorilla gorilla*, and *Pongo pygmaeus*), *Journal of Comparative Psychology*, 118 pp. 232-241.
- Call J. and Tomasello, M. 1994. Production and comprehension of referential pointing by orangutans (*Pongo pygmaeus*). *Journal of Comparative Psychology*, 108, 4 pp. 307-317.
- Call, J. and Tomasello, M. 2005. What do chimpanzees know about seeing revisited: an explanation of the third kind. Eilan, N., Hoerl, C. T. McCormack, T and Roessler, J. (eds.), *Joint Attention: Communication and Other Minds: Issues in Philosophy and Psychology* (pp. 45–64). Oxford: Oxford University Press.
- Call, J. and Tomasello, M. 2008. Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Science*, 12 pp. 187-192.
- Campbell, J. 2005. Joint attention and common knowledge. Eilan, N., Hoerl, C. T. McCormack, T and Roessler, J. (eds.), *Joint Attention: Communication and Other Minds: Issues in Philosophy and Psychology* (pp. 45–64). Oxford: Oxford University Press.

THE WORLD WE POINT AT

- Carpendale, J. I. M. and Lewis, C. 2006. *How Children Develop Social Understanding*. Oxford: Blackwell Publishers.
- Carpenter, M., Nagell, K. and Tomasello, M. 1998. Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 63, 4 pp. 1-143.
- Cattaneo L. and Rizzolatti, G. R. 2009. The mirror neuron system. *Archives of Neurology*, 66, 5 pp. 557-560.
- Clapin, H. (ed.) 2002. *Philosophy of Mental Representation*. Oxford: Clarendon Press.
- Clark, H. H. and Brennan, S. A. 1991. Grounding in communication. Resnick, L. B., Levine, J. M. & Teasley, S. D. (eds.), *Perspectives on Socially Shared Cognition* (pp. 127-149). Washington DC: APA Books.
- Cochet, H. and Vauclair, J. 2010. Pointing gestures produced by toddlers from 15 to 30 months: Different functions, hand shapes and laterality pattern. *Infant Behavior & Development*, 33 pp. 431–441.
- Csibra, G. (2005). Mirror neurons and action understanding. Is simulation involved? *Interdisciplines*.
- Csibra, G. and Gergely, G. 2006. Social learning and social cognition: the case for pedagogy. Munakata, Y. and Johnson, M. H. (eds), *Processes of Change in Brain and Cognitive Development Attention and Performance, Vol.21* (pp. 249–274). Oxford: Oxford University Press.
- Csibra, G. and Gergely, G. 2009. Natural Pedagogy. *Trends in Cognitive Sciences*, 13, 4 pp. 148-153.
- Daum, M. M., Prinz, W. and Aschersleben, G. 2008. Encoding the goal of an object-directed but uncompleted reaching action in 6- and 9-month-old infants. *Developmental Science*. 11, 4 pp. 607-619.
- Daum, M. M., Vuori, M.T., Prinz, W. and Aschersleben G. 2009. Inferring the size of a goal object from an actor's grasping movement in 6- and 9-month-old infants. *Developmental Science*, 12, 6 pp. 854-62.
- Eilan, N., Hoerlh, C., McCormack, T. and Roessler, J. 2005. (eds.), *Joint Attention: Communication and Other Minds -- Issues in Philosophy and Psychology*. Oxford: Oxford University Press.
- Eilan, N. 2005. Joint attention, communication, and mind. (2005). Joint attention, communication, and mind. Eilan, N., Hoerlh, C., McCormack, T. and Roessler, J. (eds.), *Joint Attention:*

THE WORLD WE POINT AT

- Communication and Other Minds -- Issues in Philosophy and Psychology*. Oxford: Oxford University Press.
- Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F. and Rizzolatti, G. 2005. Parietal lobe: from action organization to intention understanding. *Science*, 308 pp. 662-667.
- Fogassi, L. and Ferrari, P. F. 2007. Mirror Neurons and the Evolution of Embodied Language. *Current Directions in Psychological Science*, 16, 3. pp. 136-141.
- Franco, F. and Butterworth, G. 1996. Pointing and social awareness: declaring and requesting in the second year. *Journal of Child Language*, 23, 2 pp.307-336.
- Friesen, C. K. and Kingstone, A. 1998. The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin and Review*, 5 pp. 490-495.
- Gallese, V. and Umiltà, M.A. 2006. Cognitive continuity in primate social cognition. *Biological Theory*, 1, 1 pp. 25-30.
- Gattis, M., Bekkering, H. and Wohlschläger, A. 2002. Goal-directed imitation. Meltzoff, A. N. and Prinz, W. (eds.), *The Imitative Mind: Development, Evolution, and Brain Bases*. Cambridge: Cambridge University Press.
- Gergely, G., Bekkering, H. and Király, I. 2002. Rational imitation in preverbal infants. *Nature*, 415 p. 755.
- Gliga, T. and Csibra, G. 2009. One-year-old infants appreciate the referential nature of deictic gestures and words. *Psychological Science*, 20, 3 pp. 347-53
- Gómez, J.-C. 2007. Pointing Behaviors in Apes and Human Infants: A Balanced Interpretation. *Child Development*, 78 pp. 729-734.
- Grassmann, S. and Tomasello, M. 2010. Young children follow pointing over words in interpreting acts of reference. *Developmental Science*, 13 pp. 252-263.
- Grice, P. (1957). Meaning. *The Philosophical Review*, 64, 377-388.
- Grice, P. (1975). Logic and conversation. Cole, P. and Morgan, J. (eds.), *Syntax and Semantics*. Vol.3: Speech acts (pp.43-58). New York: Academic Press.
- Gómez, J.-C. 2007. Pointing Behaviors in Apes and Human Infants: A Balanced Interpretation. *Child Development*, 78 pp. 729-734
- Grassmann, S. and Tomasello, M. 2010. Young children follow pointing over words in interpreting acts of reference. *Developmental Science*, 13, 1 pp. 252-263.

THE WORLD WE POINT AT

- Gregory, R. L. 1997. Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society of London B*, 352 pp. 1121 - 1128.
- Grice, H. P. 1957. Meaning. *The Philosophical Review*, 64 pp. 377-388.
- Grice, H. P. 1975. Logic and conversation. P. Cole and J. Morgan, (eds.), *Syntax and Semantics*, vol. 3, Academic Press, pp. 41-58.
- Hare, B., Call, J., Agnetta, B. and Tomasello, M. 2000. Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, 59 pp. 771-785.
- Hare, B., Call, J. and Tomasello, M. 2001. Do chimpanzees know what conspecifics know? *Animal Behaviour*, 61 pp. 139-151.
- Hare, B. and Tomasello, M. 2004. Chimpanzees are more skilful in competitive than in cooperative cognitive tasks. *Animal Behaviour*, 68 pp. 571-581.
- Haye, A. A. 2008. Living being and speaking being: toward a dialogical approach to intentionality. *Integrative Psychological Behavioral Science*, 42, 2, pp. 157-163.
- Hihara S, Obayashi S, Tanaka M, and Iriki A. 2003. Rapid learning of sequential tool use by macaque monkeys. *Physiology & Behavior* 78 pp. 427-434.
- Humphrey, N. 1976. The social function of intellect. Bateson, P. P. G. and Hinde, R. A. (eds.), *Growing Points in Ethology* (pp.303-317). Cambridge, UK: Cambridge University Press.
- Hurfurd, J. R. 2003. The Neural Basis of Predicate-Argument Structure, *Behavioral and Brain Sciences*, 26, 3 pp. 261-283.
- Iacoboni, M., Woods, R.P., Brass, M., Bekkering, H., Mazziotta, J.C. and Rizzolatti, G. 1999. Cortical mechanisms of human imitation. *Science*, 286, pp. 2526-2528.
- Iverson, J. M., & Goldin-Meadow, S. 2005. Gesture paves the way for language development. *Psychological Science*, 16 pp. 367-371.
- Jackendoff, R and Landau, B. 1992. Spatial language and spatial cognition. Jackendoff, R. (ed.), *Languages Of The Mind* (pp. 99-124). Cambridge, MA: MIT Press.
- Jacob, P. 2010. Intentionality, *The Stanford Encyclopedia of Philosophy (Fall 2010 Edition)*, Zalta, E. N. (ed.), URL = <http://plato.stanford.edu/archives/fall2010/entries/intentionality/>
- Jeannerod, M. 1994. The representing brain. Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, 17 pp. 187-245.

THE WORLD WE POINT AT

- Jaswal, V. K. and Hansen, M. B. 2006. Learning words: children disregard some pragmatic information that conflicts with mutual exclusivity. *Developmental Science*, 9 pp. 158–165.
- Johnson, M. H., Mareschal, D. and Csibra, G. 2001. The functional development and integration of the dorsal and ventral visual pathways: A neurocomputational approach. Nelson, C. A. and Luciana, M. (eds.), *Handbook of Developmental Cognitive Neuroscience*. Cambridge MA: MIT Press.
- Kahneman, D., Triesman, A. and Gibbs, B.(1992). The reviewing of object files:object-specific integration of information. *Cognitive Psychology*, 24 pp. 175–219.
- Kaminski, J., Call, J. and Tomasello, M. 2008. Chimpanzees know what others know but not what they believe. *Cognition*, 109, 2 pp. 224-234.
- Kemmerer, D., Castillo J.G., Talavage, T., Patterson, S. and Wiley C. 2008. Neuroanatomical distribution of five semantic components of verbs: evidence from fMRI. *Brain and Language*, 107, 1 pp. 16-43.
- Kimura, D. 1993. *Neuromotor Mechanisms in Human Communication*. Oxford: Oxford University Press.
- Kita, S. (ed.), *Pointing: Where Language, Culture, and Cognition Meet*. Mahwah, NJ: Erlbaum.
- Knoblich, G. and Sebanz, N. 2006. The social nature of perception and action. *Current Directions in Psychological Science*, 15, 99-104.
- Kurby, C. A. and Zacks, J. M. 2007. Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, 12, 2 pp. 72-79.
- Leavens, D.A. and Hopkins, W.D. 1999. The whole hand point: The structure and function of pointing from a comparative perspective. *Journal of Comparative Psychology*, 113 pp. 417–425.
- Leavens, D. A., Hopkins, W. D. and Bard, K. A. 1996. Indexical and referential pointing in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, 110, 4 pp. 346-53.
- Leavens, D.A., Hopkins, W.D. and Bard, K.A. 2005. Understanding the point of chimpanzee pointing. Epigenesis and ecological validity. *Current Directions in Psychological Science*. 14, 4 pp. 185-189.
- Leavens, D. A., Russell, J. L. and Hopkins, W. D. 2005. Intentionality as measured in the persistence and elaboration of communication by chimpanzees (*Pan troglodytes*). *Child Development*, 76 pp. 291–306.

THE WORLD WE POINT AT

- Leung, E. H. L., and Rheingold, H. L. 1981. Development of pointing as a social gesture. *Developmental Psychology*, 17, 2 pp. 215–220.
- Liebal, K., Carpenter, M. and Tomasello, M. 2010. Infants' use of shared experience in declarative pointing. *Infancy*, 15, 5 pp. 545-556.
- Liszkowski, U. 2008. Before L1: A differentiated perspective on infant gestures. *Gesture*, 8, 2 pp. 180-196.
- Liszkowski, U., Carpenter, M., Henning, A., Striano, T. and Tomasello, M. 2004. Twelve-month-olds point to share attention and interest. *Developmental Science*, 7 pp. 297-307.
- Liszkowski, U., Carpenter, M., Henning, A., Striano, T. and Tomasello, M. 2006. Twelve- and 18-month-olds point to provide information for others. *Journal of Cognitive Development*, 7 pp. 183-187.
- Milner, A. D. and Goodale, M. A. 2008. Two visual systems re-viewed. *Neuropsychologia*, 46 pp. 774–785.
- Moll, H. and Tomasello, M. 2007. Cooperation and human cognition: The Vygotskian Intelligence Hypothesis. *Philosophical Transactions of the Royal Society B*, 362, 1480 pp. 639–648.
- Moll, H., Carpenter, M. and Tomasello M. 2007. Fourteen-month-olds know what others experience only in joint engagement. *Developmental Science*, 10, 6 pp. 826-35.
- Molenberghs, P., Cunnington, R. and Mattingley, J. 2009. Is the mirror neuron system involved in imitation? A short review and meta-analysis. *Neuroscience and Biobehavioral Reviews*, 33, 7 pp. 975-980.
- Moore, C. & Dunham, P. J. 1995. *Joint Attention: Its Origins and Role in Development*. Lawrence Erlbaum.
- Mundy, P., Block, J., Vaughan Van Hecke, A., Delgado, C., Parlade, M. and Pomeroy, Y. 2007. Individual differences in the development of joint attention in infancy. *Child Development*, 78 pp. 938–954.
- Mundy, P. and Newell, L. 2007. Attention, Joint Attention, and Social Cognition. *Current Directions in Psychological Science*, 16, 5 pp. 269-274.
- Newman-Norlund, R. D., van Schie H. T., van Zuijlen A. M. and Bekkering, H. 2007. The mirror neuron system is more active during complementary compared with imitative action. *Nature Neuroscience*, 10, 7 pp. 817-818.

THE WORLD WE POINT AT

- Nowak, M. A. 2006. Five rules for the evolution of cooperation, *Science*, 314, 5805, pp. 1560–63.
- Parsons, T. (1990). *Events in the Semantics of English*. Cambridge: MIT Press.
- Penn, D. C., Holyoak, K. J. and Povinelli, D. J. 2008. Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences*, 31, 2 pp. 109-178.
- Plotkin H. 2003. We-intentionality: an essential element in understanding human culture? *Perspectives in Biology and Medicine*, 46, 2 pp. 283-296.
- Pollick, A. S. and de Waal, F. B. M. 2007. Ape gestures and language evolution. *Proceedings of the National Academy of Sciences*, 104, 19 pp. 8184-818
- Povinelli, D. J. 2001. The minds of humans and apes are different outcomes of an evolutionary experiment. S. Fitzpatrick and J. Bruer (eds.), *Carving Our Destiny: Scientific Research Faces a New Millennium*. pp. 1–40. Washington, D.C.: National Academy of Sciences and Joseph Henry Press.
- Povinelli, D.J. 2004. Behind the ape's appearance: Escaping anthropomorphism in the study of other minds. *Daedalus: Journal of the American Academy of Arts and Sciences* , Winter, pp. 29-41.
- Povinelli, D.J. and Vonk, J. 2003. Chimpanzee minds: Suspiciously human? *Trends in Cognitive Science*, 7 pp. 157-160.
- Povinelli, D.J., Bering, J. and Giambrone, S. 2003. Chimpanzee 'pointing': Another error of the argument by analogy? S. Kita (ed.), *Pointing: Where Language, Culture, and Cognition Meet* (pp. 35–68). Hillsdale, NJ: Erlbaum.
- Povinelli, D. J., Theall, L. A., Reaux, J.E. and Dunphy-Lelii, S. 2003b. Chimpanzees spontaneously alter the location of their gestures to match the attentional orientation of others. *Animal Behaviour*, 66 pp. 71-79.
- Proffitt, D. R. 2006. Distance Perception. *Current Directions in Psychological Science*, 15, 3 pp. 131-135.
- Pylyshyn, Z. W. 2000. Situating vision in the world. *Trends in Cognitive Sciences*, 4, 5 pp. 197-207.
- Rao, S. C., Rainer, G. and Miller, E. K. 1997. Integration of What and Where in the Primate Prefrontal Cortex. *Science*. 276, 5313 pp. 821-824.
- Rizzolatti, G., & Arbib, M. A. 1998. Language within our grasp. *Trends In Neurosciences*, 21, 188–194.
- Rizzolatti, G. and Craighero, L. 2004. The mirror-neuron system. *Annual Review of Neuroscience*, 27 pp. 169–192.

THE WORLD WE POINT AT

- Rizzolatti, G., Fogassi, L. and Gallese V. 2001. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2 pp. 661–70.
- Saylor, M., Baldwin, D. and Sabbagh, M. 2004. Converging on word meaning. Hall, D. G. and Waxman, S. R. (eds.), *Weaving a Lexicon* (pp. 509–531). Cambridge, MA: MIT Press.
- Sebanz, N., Bekkering, H. and Knoblich, G. 2006. Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10 pp. 70-76.
- Sebanz, N., Knoblich, G. and Prinz, W. 2003. Representing others' actions: Just like one's own? *Cognition*, 88 pp. B11–B21.
- Samson, D. and Apperly, I. A. 2010. There is more to mind reading than having theory of mind concepts: new directions in theory of mind research. *Infant and Child Development*, 19, 5 pp. 443-454.
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J. and BodleyScott, S. E. In press. Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*.
- Schneider, W. and Shiffrin, R. M. 1977. Controlled and automatic human information processing. I. Detection, search, and attention. *Psychological Review*, 84, 1 pp. 1–66.
- Searle, J. R. 1983. *Intentionality*. Cambridge: Cambridge University Press.
- Searle, J. R. 1995. *The Construction of Social Reality*. New York: Free Press.
- Sahin, E. and Erdogan, S. T. 2009. Towards linking affordances with mirror/canonical neurons. *Proceedings of 24th International Symposium. of Computer and Information Sciences*, pp. 397-404.
- Schwe, H., and E. Markman. 1997. Young children's appreciation of the mental impact of their communicative signals. *Developmental Psychology*, 33 pp. 630–635.
- Searle, J. R. 1995. *The Construction of Social Reality*. New York: Free Press.
- Senju, A., Csibra, G. and Johnson, M. H. 2008. Understanding the referential nature of looking: infants' preference for object-directed gaze. *Cognition*, 108 pp. 303–319.
- Shepherd, S. V., Klein, J. T., Deaner, R. O., and Platt, M. L. 2009. Mirroring of attention by neurons in macaque parietal cortex. *Proceedings of the National Academy of Sciences U.S.A.*, 106, 23 pp. 9489–9494.

THE WORLD WE POINT AT

- Smith, E. E. and Kosslyn, S. M. 2007. *Cognitive Psychology: Mind and Brain*. New Jersey: Pearson/Prentice Hall
- Soames, S. 2010. *What is Meaning?* Princeton: Princeton University Press.
- Soames, S. In press. Propositions. Fara, D. G. and Russell, G. (eds.), *The Routledge Companion to the Philosophy of Language*. New York: Routledge.
- Southgate, V., Chevallier, C., and Csibra, G. 2009. Sensitivity to communicative relevance tells young children what to imitate. *Developmental Science*, 12 pp. 1013-1019.
- Southgate, V., Senju, A. and Csibra, G. 2007. Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, 18 pp. 587-592.
- Sperber, D. and Wilson, D. 1986. *Relevance: Communication and Cognition*. Oxford: Blackwell.
- Striano, T., Chen, X., Cleveland, A. and Bradshaw, S. 2006. Joint attention social cues influence infant learning. *European Journal of Developmental Psychology*, 3, 3 pp. 289-299.
- Tagliatela, J. P., Russell, J. L., Schaeffer, J. A. and Hopkins, W. D. 2008. Communicative signaling activates 'Broca's' homolog in chimpanzees. *Current Biology*, 18, 5 pp.343–348.
- Tomasello, M. 1999. *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.
- Tomasello, M. 2001. Perceiving intentions and learning words in the second year of life. M. Bowerman and S. Levinson (eds.), *Language Acquisition and Conceptual Development* (pp. 132–158). Cambridge: Cambridge University Press.
- Tomasello, M. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge: Harvard University Press.
- Tomasello, M. 2006. Why don't apes point? Enfield, N. and Levinson, S. (eds.), *Roots of Human Sociality* (pp. 506-524), New York: Wenner-Grenn.
- Tomasello, M. 2008. *Origins of Human Communication*. Cambridge, MA: MIT Press.
- Tomasello, M. 2009. *Why We Cooperate*. Cambridge: Cambridge, MA: MIT Press.
- Tomasello, M. and Abbot-Smith, K. 2002. A tale of two theories: response to Fisher. *Cognition*, 83 pp. 207–214.
- Tomasello, M., Call, J. and Hare, B. 2003. Chimpanzees understand the psychological states of others – the question is which ones and to what extent. *Trends in Cognitive Sciences*, 7 pp. 153-157.

THE WORLD WE POINT AT

- Tomasello, M. and Carpenter, M. 2005. The emergence of social cognition in three young chimpanzees. *Monographs of the Society for Research in Child Development*, 70, 1 pp. vii-132.
- Tomasello, M. and Carpenter, M. 2007. Shared intentionality. *Developmental Science*. 10, 1 pp. 121-125.
- Tomasello, M., Carpenter, M., Call, J., Behne, T. and Moll, H. 2005. Understanding and sharing intentions: The ontogeny and phylogeny of cultural cognition. *Behavioral and Brain Sciences*, 28 pp. 675-735.
- Tomasello, M., Carpenter, M. and Liszkowski, U. 2007. A new look at infant pointing. *Child Development*, 78 pp. 705-722.
- Tomasello, M. and Haberl, K. 2003 Understanding attention: 12- and 18-month-olds know what is new for other persons. *Developmental Psychology*, 39 pp. 906–912.
- Tomasello, M. and Rakoczy, H. 2003. What makes human cognition unique? From individual to shared to collective intentionality. *Mind & Language*, 18, 2 pp. 121-147.
- Triesch, J., Teuscher, C., Deak, G.O. and Carlson, E. 2006. Gaze following: Why (not) learn it? *Developmental Science*, 9 pp. 125–14.
- Tuomela, R. 2002. Collective goals and communicative action. *Journal of Philosophical Research*, 27 pp. 29–64.
- Van Essen, D. C., Anderson, C. H. and Felleman, D. J. 1992. Information processing in the primate visual system: an integrated systems perspective. *Science*, 255 pp: 419-423.
- van der Wel, R. P. R. D., Sebanz, N. and Knoblich, G. In press. Action perception from a common coding perspective. Johnson, K. and Schiffrar, M. (eds.), *Visual Perception of the Human Body in Motion*.
- Vauclair, J., & Imbault, J. 2009. Relationships between manual preferences for manipulative actions and for pointing gestures in infants and toddlers. *Developmental Science*, 12, 1060–1069.
- Volterra, V., Caselli, M.C., Capirci, O. and Pizzuto, E. 2005. Gesture and the emergence and development of language. M. Tomasello and D. Slobin (eds.), *Beyond Nature-Nurture. Essays in Honor of Elizabeth Bates* (pp. 3-40). N.J.: Lawrence Erlbaum Associates.
- Whiten, A., Horner, V. and de Waal, F. B. M. 2005. Conformity to cultural norms of tool use in chimpanzees. *Nature*, 437 pp. 737-740.

THE WORLD WE POINT AT

- Xu, F. 1999. Object individuation and object identity in infancy: The role of spatiotemporal information, object property information, and language. *Acta Psychologica*, 102 pp. 113-136.
- Yoon, J. M. D., Johnson, M. H. and Csibra, G. 2008. Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences U.S.A.*, 105 pp. 13690–13695