The Evolution of Joint Attention: A Review and Critique

Timothy P. Racine, Tyler J. Wereha, Olga Vasileva, Donna Tafreshi and Joseph J. Thompson

Abstract Joint attention can be defined as the ability to intentionally coordinate an attentional focus on some object or state of affairs with another. This capacity is believed by most theorists to be logically, developmentally, and evolutionarily prior to language and further forms of social cognition tied up with human social communication. However, although there has been a good deal of empirical and theoretical work on joint attention, there has been less attention paid to the evolution of joint attention in its own right. There has also been sustained debate concerning whether other primates can be said to engage in joint attention, which in turn conditions the evolutionary theories that are offered. In this chapter, we define and describe joint attention, discuss the skills it involves, and the extent to which we share these with other animals. Next, we review work that has been done on the evolution of joint attention and related capacities and classify it as a function of its mode of explanation. We then discuss the aforementioned forms of evolutionary explanation in the light of recent evolutionary theories and findings that question adaptationist thinking, and consider the potential relevance of non-adaptationist thinking for theoretical work on the evolution of joint attention.

Keywords Joint attention • Aadaptationist stance • Evolutionary psychology • Core knowledge • Shared intentionality • Non-human primates

Perhaps unsurprisingly, providing a satisfactory and relatively complete account of the evolution of social communicative abilities in human and non-human primates has proven to be a difficult task. Although it is important to understand the origins and properties of any animal communication system as an end in itself, interest

T. P. Racine (✉) · T. J. Wereha · O. Vasileva · D. Tafreshi · J. J. Thompson
Department of Psychology, Simon Fraser University, 8888 University Drive,
Burnaby, BC V5A 1S6, Canada
e-mail: tracine@sfu.ca

M. Pina and N. Gontier (eds.), *The Evolution of Social Communication in Primates*,
Interdisciplinary Evolution Research 1, DOI: 10.1007/978-3-319-02669-5_7,

in evolutionary analyses of communicative systems that appear similar to those in humans is almost universal. That is, whereas a non-human organism does not need a full-fledged language, and the biological, sensorimotor, and cognitive abilities that human languages seem to require, in order for "social communication" to be possible, such communication exhibits near universal interest to the extent that it can reveal something about the common descent, or perhaps convergent evolution, of the skills under investigation.

Part of the reason for the difficulty of this task is that there might be, pace Darwin, qualitative rather than quantitative differences in linguistic skill and repertoire between humans and our nearest living relative, the common chimpanzee, Pan troglodytes, or indeed the chimpanzee's nearest living relative, the bonobo, Pan paniscus (see Tallis 2011 for a recent forceful expression of one such view, see also e.g. Penn et al. 2008). Another reason is that natural selection itself might not be the right or at least sole theoretical tool required to understand the evolutionary origins of this particular phenomenon for the simple reason that "social communication" might not arise from a simple adaptation or set of adaptations (Gould and Lewontin 1979; Lickliter and Honeycutt 2013; Racine 2013). In this chapter, we first briefly introduce the evolution of social communication in general terms (Sect. 1) to justify focussing on a particular capacity called "joint attention" that most agree is required for language and intentional social communication in general. In Sect. 2, we define and describe joint attention further, discuss the skills it involves, and the extent to which we share these with other animals. In Sect. 3, we review work that has been done on the evolution of joint attention and related capacities and classify it with respect to its degree of reliance on adaptationist and innatist thinking. In Sect. 4, we conclude by discussing the aforementioned forms of evolutionary explanation in the light of recent evolutionary theories and findings that question adaptationist thinking, and consider the potential relevance of non-adaptationist thinking for theoretical work on the evolution of joint attention. In particular, we discuss the relation between development and evolution and their potential reintegration.

1 Introduction: Language, Social Communication, and Joint Attention

Even if, as Darwin would have it, there is continuity rather than discontinuity in the social communication systems of humans and other primates, it has proven difficult, at least if understood in selectionistic terms, to explain the evolution of language through a series of gradual small steps. Darwin's (1879) own view, laid out in The Descent of Man, emphasizes changes in cognition and general intelligence followed by sexual selection acting on vocal control. Through this process, a "musilanguage", as Brown (1999) calls it, is created that adds meaning to these voicings, which in turn is the result of, and leads to, changes in general intelligence. Although few, including Darwin, argue for direct selection on a language

organ or instinct (cf. Pinker 1994), this has proven to be an area of evolutionary theory that is quite resistant to significant progress or at least consensus. Part of the reason for this is probably that the concept "language" itself is thorny and used in a variety of ways, each of which potentially requires slightly different skills. There is also a large and somewhat unwieldy literature in the philosophy of language, philosophy of mind, and philosophy of biology, that is, relevant for this discussion, but is difficult to incorporate in a sustained manner.

This is not to say though that there have not been many attempts. Although several others could be listed, for comparative purposes, let us consider three very different recent works in this vein. Hauser et al. (2002), Place (2000), and Whiten and Erdal (2012) account for language evolution chiefly by emphasizing the evolution of a particular mental-combinatorial skill (recursion), manual and representational skills (including pointing), and sociocultural skills (mostly mindreading abilities), respectively. Given that there is little agreement in explanatory scope or focus in these sorts of "grand theory" attempts, it may make some sense to focus on a thorough explanation of a particular aspect of language that make it possible. This would mean resisting the urge to single out one factor that one theorizes to be chiefly responsible for human language, such as recursion in the case of Hauser et al. (2002). For this reason, we are more in sympathy with Whiten and Erdal's (2012) attempt at identifying a potential suite of such factors. However, the fact that Whiten and Erdal (2012: 2127) added a footnote at the proof stage in response to an article on the "cultural niche" by Boyd et al. (2011) suggests that Whiten and Erdal realize their notion of a "(socio-)cognitive niche" might be associated with the Evolutionary Psychology theorizing of Pinker (2010), an association that Whiten and Erdal are clearly attempting to avoid. Furthermore, from our perspective and probably Boyd et al. (2011) as well, Whiten and Erdal (2012) might be assuming too much adaptive value in "mindreading" skills in an ancestral context.

We think we can make some headway on these issues by focussing on a capacity that most agree is required for language and intentional social communication in general. This capacity is joint attention, which Leaven and Racine (2009: 240) define as "the ability to intentionally co-orient towards a common focus". However, as with language and social communication more broadly, joint attention may not be a single capacity, and depending on the researcher, can involve gaze following of various forms, social referencing, gesture including different forms of pointing gesture, and instrumental imitation (for a review, see Racine and Carpendale 2007). Joint attention is also used in the field in an inclusive manner to refer to this entire suite of behaviours (or some subset of them), or an exclusive manner to mean literal episodes of joint (typically visual) attention. However, if joint attention is necessary for language, it is therefore logically and evolutionarily prior to it, and is therefore an ability that needs to be explained in a satisfactory evolutionary account of social communication.

One can also easily see why such a skill would be implicated in language and intentional social communication: it would seem that minimally an individual has to appreciate that another is referring to something in order to learn the meaning of words or to intentionally inform another of some state of affairs (but see Akhtar

and Gernsbacher 2007).[1] Although this will not be the focus of our chapter, intentional social communication would likely necessarily involve joint attention as well, although there might be classes of social communication (e.g. alarm calls) that may not be intentional, at least not in the same sense and therefore do not qualify as involving joint attention in the way that it is used in the research community. In either case, although joint attention is a relatively straightforward ability, we need to briefly acknowledge the complexity of the joint attention conceptual domain before proceeding. This is because this complexity has implications for theories concerning the evolution of joint attention, and also explains why in our review we focus on a variety of accounts, some of which are broader classes of theory.

1.1 Joint Attention and a Motley Crew of Related Social Cognitive Concepts

Trevarthen (1979) and Trevarthen and Hubley (1978) classified two important developmental transitions in human infants in the first year of life that he termed primary and secondary intersubjectivity, respectively. The first phase refers to a shared awareness shown in social games, turn taking, and emotional exchanges with a caregiver that is evident in the social smiles that begin around 2 months. This early period is dyadic—back and forth between infant and caregiver, but does not include an additional common object of focus. In a semantic idiom, the interaction lacks extension. The absence of a common focus precludes there being referent for the interaction and, in this sense, the interaction is not about anything. This is in contrast to secondary intersubjectivity that involves a common focus on some object or state of affairs. What psychologists, philosophers, linguists, and cognitive and brain scientists call "joint attention" is in many respects the same phenomenon that Trevarthen terms secondary intersubjectivity. Other than reflecting or perhaps enabling secondary intersubjectivity, joint attention is also viewed as an early developing form of a general "mindreading" ability, and discussed as an aspect of metacognition, metarepresentation, and/or "theory of mind" (Brinck and Liljenfors 2013; Call and Tomasello 2008; Carruthers 2009). Although these are all somewhat related social cognitive domains, part of the difficulty in explaining the evolutionary origins of joint attention is that this concept is tied up with a variety of other social cognitive ones.

[1] Although Akhtar and Gernsbacher (2007) argue for robust effects of "overhearing" in typical language development, the concern in the present paper is with the evolutionary origins of social communication. In this context, it is difficult to imagine a scenario where language or protolanguage could evolve independently of joint attention. However, if Akhtar and Gernsbacher (2007) are reacting to the often highly cognitive descriptions of language that abound in the developmental literature, we share their concern.

It might be tempting to unite these various social cognitive concepts by speaking of early versus later developing social cognitive skills, and therefore of joint attention in contradistinction to the understanding of more complex states, such as beliefs, generally believed to be typical of preschoolers. However, as we will see, theorists such as Baillargeon et al. (2010) have used looking time studies to argue for belief understanding in infants. Although others such as Charles and Rivera (2009) argue that infant looking time procedures are more appropriately interpreted as measures of infant perception and object-oriented behaviours rather than measures of infant knowledge, Luo (2011), for example, adopts the latter interpretation and claim that belief understanding, and therefore metarepresentation, is present as early as 10 months of age. This is the same time period as Trevarthen's secondary intersubjectivity, and the rudiments of joint attentional skills like pointing and gaze following. Thus, although we focus on joint attention in what follows, we need to return to these more general concepts when discussing evolutionary theories of joint attention. This is in part because there is little evolutionary work on discrete joint attention behaviours (e.g. pointing and gaze following), and also because its role in further social cognitive and communicative development is often situated in these broader social cognitive accounts.

2 Joint Attention: Continuity or Discontinuity?

Joint attention would seem to be a capacity, that is, evolutionarily speaking, quite ancient.[2] After all, it is the ability to coordinate one's another focus with that of a conspecific ("to jointly attend"). It also should be obvious when two individuals of a given species are or are not engaged in an act of joint attention. However, a complication is that two individuals might be looking at the same thing by happenstance without an awareness of the attentional focus of the other jointly attending individual. Perhaps though this should not necessarily mean that it does not count in some sense as joint attention, for example Butterworth (1998: 171) defines joint attention as "simply…looking where someone else is looking". Similarly, Bakeman and Adamson (1984) allow for a category they term "passive" joint attention. However, the way the concept is typically used in the field denotes a mutual awareness of each other's attentional state. To put this differently, the way "joint attention" is typically understood within the field, both parties must intend to jointly attend to some shared object or state of affairs, and be aware that the other's attention is also focussed on the same object or state of affairs (Leavens and

[2] Given that the original discovery of mirror neurons some 20 years ago was in macaques, it would seem that a capacity for joint attention might have its roots as far back as old world monkeys. However, it is not clear what mirror neurons really do, or the extent to which they imply "mindreading" abilities like joint attention (Racine et al. 2012). For one thing, although all species of great ape pass mirror self-recognition tests, monkeys do not.

Racine 2009). That is, they must engage in what is often called triadic interaction (e.g. Tomasello 1995). As long as one keeps in mind that this just means that the attribution of joint attention to two agents involved in the right sort of activity (e.g. gaze following, pointing, and so forth) requires that their coordinated mutual attentional focus be non-accidental ("that they intend to attend"), then this is a reasonable terminological stipulation. And, on the face of it, one would still expect joint attention to be a capacity shared with some other primate species through common descent, and probably other highly social species through convergent evolution, as has been suggested, for example, in work on the corvid, Corvus corax, the common raven (Pika and Bugnyar 2011).

It might be surprising, then, for the reader without a previous background in this research area to learn that there is a considerable amount of pessimism concerning the mentality of non-human animals when it comes to explaining basic joint attentional capacities. This is not the place to review the causes and consequences of this state of affairs, but rather to see how they play out in evolutionary work on joint attention. However, a summary of some of the key concerns is required in order to proceed (for more detail, see, e.g. Racine 2012a, b).

Scepticism regarding non-human joint attentional capacities often involves two steps of argument, the first of which is conceptualizing the triad in "triadic interaction" not as "two agents coordinating their attention to an object", but rather as two "coordinated subject–object mental relations". That is, the activity of joint attention has been redescribed as a shared second-order representational state.[3] As long as it is understood that this is just to say there is a mutual awareness of an intention to attend, this is not inherently problematic. However, the second-order mental states in question are not typically understood in this manner, but rather causally. That is, the mental states are theorized to give rise to the behaviour of interest rather than the joint attentional behaviour being grounds for the attribution of the mental state. The underlying reason for conceiving as joint attention as mental representations of subject–object relations—and causal ones at that—is that joint attention is typically understood from within the lens of what is often called in the philosophy and cognitive sciences, the representational theory of mind (RTM) (see Racine 2012b; Slaney and Racine 2011). The next step, of course, is to discount activities that can be given ostensibly "less mental" interpretations. And the pessimism about animal minds comes full circle.

A highly influential group of such pessimists is Tomasello and colleagues who have consistently asserted that shared attention involving great apes such as chimpanzees should not qualify as joint attention in the way used in the field (e.g. Tomasello et al. 2005, 2007; Tomasello 2008, 2014). The evolutionary force

[3] See Andrews (2012) and Hutto and Myin (2013) for discussion and critiques of the more general position that folk psychology implies the uncovering of propositional attitudes. Alternative approaches maintain that "the folk" understand one another as entire organisms with histories, embedded in particular contexts, with moods, temperaments, and so on, not typically as bearers of propositional attitudes.

of their argument will be considered in more detail below (see also Racine 2012a, 2013; Wereha and Racine 2012), but the most relevant point, for them, is that apes' motivations for sharing attention are mostly instrumental, whereas humans do the same activities for non-instrumental purposes (to simply share intentions as an end in itself). This follows a classical distinction between prelinguistic imperative and declarative acts (so-called protoimperatives and protodeclaratives) introduced by Bates et al. (1975). But, unlike Bates et al., Tomasello et al. conflate the cause of the behaviour with the meaning of the behaviour. This fundamental difference plays out in Tomasello's evolutionary work on joint attention and related work by others.

Consider, by parity, the following. Assume that in an act of shared attention one organism is highly motivated to attend because of their interest in the interaction, but the other organism is only doing so begrudgingly, or even for a completely different reason (e.g. they were paid, or otherwise reinforced, for doing so), but they were still both intentionally jointly attending. Would we be justified in saying that therefore only one of the organisms is really sharing attention (or really pointing cf. Tomasello 2006)? We might say one was more motivated than the other, and it might well be that the more motivated one was more likely to initiate or continue episodes of joint attention, but it is an obvious tautology that this would not and could not mean that only one of the two was jointly attending. Therefore, to claim that only one organism is really jointly attending, as stated, is to conflate the cause and meaning of the activity. To claim that both are attending because of some common underlying neural causes is, of course, equally problematic. The typical person lacks knowledge of the neural causes of behaviour, yet still has sensible grounds for describing the activity correctly.

We assume that many readers will agree that substituting a motivational construct for an intentional–attentional one is scientifically very confusing and presumably ill-advised. However, it is this very move that Tomasello (2008, 2014) performs in his "shared intentionality hypothesis", which is more an illustrative use of a body of work in the philosophy of action than a scientific hypothesis (see Racine 2012b). By contrast, Leavens and Racine (2009) concluded that while there is joint attentional variation among apes that is attributable to differential rearing histories, apes engage in all behaviours considered to show joint attention in humans.[4] Although we will not repeat this point again in the present chapter or include it in our classification of theories, Leavens and Racine (2009) argue that therefore there is no clear evidence for uniquely human cognitive adaptation(s) for joint attention.

Even if Leavens and Racine are wrong, it is important to bear in mind that claiming that chimpanzees are aware of the attention of others and accordingly intentionally engage in joint attention when they follow gaze or gesture is not

[4] In terms of relevant similarities, it seems noteworthy that all great ape species can pass any test of gaze following ability that a human child can, including being aware that individuals cannot see though obstructions, and also that great apes use manual gestures, including pointing gestures, particularly in captive environments (Brauer et al. 2005; Leavens and Racine 2009; Pika 2008).

tantamount to claiming that a chimpanzee would have to be aware of everything about another's attentional focus. Or, that an ape would need to understand all the subtle ways that attention can be deployed (Wilkins 2003). Or, even that the identical genetic, neural, sensorimotor, cognitive, or behavioural substrates would need to be responsible for joint attention in humans and other apes.[5] After all, the same would apply if comparing a prelinguistic human to a human child or adult. But this is not the question we asked to begin with. We simply asked whether an ape (or a prelinguistic infant as the case may be) intentionally shares in the attentional focus of a conspecific around some object of mutual attention and is aware that his or her interlocutor is doing the same. To forget this would again invite potential confusion of definitional issues of what counts as joint attention and empirical (and sometimes causal) issues concerning how joint attention occurs (Racine 2012b).

Although many researchers seem to assume that these matters can be resolved empirically or by creating more adequate methodologies, the foregoing should suggest these are not simple empirical or methodological matters.[6] That is, one cannot stipulate that other animals do not have a capacity for joint attention just because one believes that their joint attentional behavioural can be explained without the attribution of second-order representational states—particularly when there is no clear evidence that human infants require second-order representational states to engage in joint attention (Leavens 2012; Leavens and Racine 2009; Racine 2012b; Racine et al. 2012). After these considerations, we are now, we believe, in a better position to review and discuss the evolution of joint attention.

3 Evolutionary Theories of Joint Attention

Evolutionary claims about joint attention follow the tendency in the field to conceive of a joint attention in an inclusive or exclusive manner. It is also common to focus on a particular capacity, such as pointing or gaze following. Pointing in particular is thought to have played an important role in human evolution, being a human specific adaptation in that it is considered to be a human universal (Butterworth 2003; Povinelli et al. 2003). And, it is argued to be absent among wild populations of great apes (Tomasello 2006, 2008). The pointing gesture itself has also been considered to be an adaptation of human physiology (Butterworth

[5] Although this might seem bad news for an attempted evolutionary analysis of joint attention, it might on the other hand be diagnostic of the limitations of adaptationist and selectionist thinking. An important change in thinking in the past 20 years or so is to question the opposition between evolution and development (e.g., Gottlieb 2002; Jablonka and Lamb 2005; Lickliter and Honeycutt 2013; Oyama 2000).

[6] This is not to say though that empirical matters are independent of these sorts of conceptual concerns (Glock 2013). The point is that improved methods or additional empirical work cannot avoid conceptual problems that result for the application of the RTM framework when the framework itself is built into the interpretation of the new methods or findings.

2003) and Povinelli and Davis (1994) have documented differences in the resting state of the index finger in humans and chimpanzees, suggesting biological preparation in humans.

The focus on the evolution of pointing is not surprising because pointing is taken to be the least ambiguous indicator of shared attention skill in that it inherently requires another and some shared object or state of affairs. In a strictly behavioural sense, pointing is a very useful indicator because it is more overt than gaze following and the latter develops earlier than pointing when its function is less clear. However, gaze following often develops into a form that involves so-called visual checking, wherein one individual double-checks to ensure that the attended visual referent was the intended one (Morissette et al. 1995), which seems as cognitively demanding in a joint attentional sense. Also, as summarized earlier, there is quite a bit of debate about whether pointing requires second-order representation, and therefore, whether it should be seen as a "superior" indicator of joint attentional skill in a logical sense. Therefore, a behaviour-by-behaviour comparison of evolutionary joint attention claims seems ill-advised, and we will present the theories according to their form of evolutionary theory. As we will see, the three approaches we compare are all explicitly adaptationist. Try as we might, we were also not able to find other forms of explanation used for joint attention.

We will describe the theories in broad strokes and essentially outline what are common but potentially problematic ways to think about the evolution of joint attention. In this sense, we are not concerned with giving the high-resolution details of a given account, and suggest, if desired, that interested readers consult the accounts in question for more detail. We turn to possible remedies in the following section. We begin with an orienting discussion of Trevarthen's work on secondary intersubjectivity to show the difficulty in making an evolutionary explanation of an innatist claim. We then proceed to review three main ways of thinking about the evolution of joint attention beginning with what is the most extremely adaptationist, and to our mind, straightforwardly problematic account, which is that of evolutionary psychology. We then consider the core knowledge account of social cognitive abilities, which are argued to be present at the same time that joint attentional skills emerge. Here, the emphasis switches from adaptationism to innatism, but the common thread is domain-specificity. We conclude with Tomasello's shared intentionality theory, which is often understood to be a more moderate alternative to evolutionary psychology and core knowledge accounts, despite the fact that it contains many of the same problematic assumptions.

3.1 Intersubjectivity and the Perils of Innatist Explanation

Trevarthen's innatist explanation of secondary intersubjectivity anticipates in some respects accounts as diverse as core knowledge theory in the sense of the implicit Chomskyan "poverty of the stimulus" flavour to the argument, and Tomasello's shared intentionality theory. Trevarthen unabashedly claims that the capacity for

intersubjectivity, and by extension, joint attention, is innate. Although we do not have the space to review his theory and the support for it in detail, he recently summarized his position, in what in many respects is a reflection on his distinguished career (Trevarthen 2011: 119), in the following manner: "We are born to generate shifting states of self-awareness, to show them to other persons, and to provoke interest and affectionate responses from them", and "cultural intelligence itself is motivated at every stage by the kind of powers of innate intersubjective sympathy that an alert infant can show shortly after birth". Beyond Trevarthen's landmark infant development studies which demonstrated what he called protoconversations between infant and caregivers and coupling between infant gestures and prespeech, Trevarthen's theory is supported to some extent by Meltzoff's work on neonatal imitation (but see e.g. Jones 2009), and the discovery of mirror neurons which have been argued by some to underlie intersubjectivity (e.g. Gallese 2001, but see e.g. Racine et al. 2012). However, the chief impetus for this theory, we think, is more intuitive.

Trevarthen often criticizes the information processing and overly cognitive manner is which infant life is—or at least was—often explained. Like Bruner (e.g. 1983), who cites Trevarthen's work as support for his own, the goal is to emphasize sociocultural processes that are presupposed in mainstream accounts, in Bruner's case behaviourism (Racine 2012b), and in Trevarthen's Piagetian accounts. In many respects, we are in sympathy with the motivations of Trevarthen and Bruner—and also Tomasello, another prominent theorist who emphasizes sociocultural factors; information processing and unqualified learning or constructivist theories are too simple at worst and redescriptions of familiar phenomena at best to be of much theoretical use in a general sense. However, claiming that something is innate or the product of natural selection (i.e. an adaptation) is not necessarily that much of an improvement. That is, although we do understand the limitations of the unqualified "socioculturalism" or "interactionism" that, for example, Tomasello has criticized in cultural theories (Wereha and Racine 2012), claiming that something is innate or the product of an adaptation tells us as much about the development of psychological capacities as cultural notions of appropriation and internalization do. That is, very little.

3.2 Evolution Psychology: Massively Modular and Massively Ambitious

Evolutionary psychologists, by which we mean to single out the so-called Santa Barbara school of Evolutionary Psychology (EP) associated with the work of Tooby and Cosmides, Buss, Pinker, and a few others, have also gained ascendance within psychology in the past 25 years for reasons that are ironically very similar to those of Trevarthen, Bruner, and Tomasello. Although EP takes this to fanciful levels by proposing that psychology itself should come home to EP as a core explanatory principle, in criticizing what EP has called the standard social science

model, they are essentially taking the field to task for neglecting the biological substrate of cultural learning. Tooby and Cosmides (1992) and Pinker (1994) in particular have traded on the notion that psychologists too easily fall back on naive forms of learning theory to explain human capacities. This critique is well taken as far as it goes, but of course the remedy suggested by EP—namely that of adopting EP—has been thoroughly criticized on a number of fronts, including its fundamental conceptual and biological adequacy (see e.g. Heyes 2012; Lickliter and Honeycutt 2013; Racine 2013; Wereha and Racine 2012). As this topic has been covered very well in the literature, we will not repeat these critiques here, but will focus on what is more relevant for present concerns. The evolutionary foundations of core knowledge and shared intentionality theory have not received nearly the same level of scrutiny, and we therefore focus more attention on those sections.

According to EP, joint attention and other aspects of social cognition are adaptations in a classical sense. In particular, they are adaptations to selective pressures our hominid ancestors faced in the Pleistocene. As Tooby and Cosmides, the founders of the approach, write in the foreword of Baron-Cohen's (1995) book Mindblindness: An Essay on Autism and Theory of Mind, "our cognitive architecture resembles a confederation of hundreds or thousands of functionally dedicated computers (often called modules) designed to solve adaptive problems endemic to our hunter-gatherer ancestors" (p. xiii). Baron-Cohen's (1995) "human mindreading system" is his explicit contribution to the EP programme of research.[7] Within this system, he describes four major modules that process social information that he refers to as the intentionality detector (ID), the eye-direction detector (EDD), the shared attention mechanism (SAM), and the theory of mind mechanism (ToMM). These modules process information on agents' volition, perception, shared attention, and epistemic states, respectively. This model has been persuasive in regards in its proposed aetiology of autism as damage to the SAM or ToMM, according to the model, leads to the kind of social cognitive impairments that presents in autism spectrum disorder (ASD).[8] From this perspective, although Baron-Cohen claims the ID and EDD are shared with other primates, a specific adaptation, the SAM, does the heavy lifting in explaining the existence of human joint attention capacities. True to their adaptationist programme, Tooby and Cosmides (1995: xvi) contend that, "natural selection is the only known natural process that builds functional organization into the species-typical designs of organisms". Be that as it may, this reasoning has failed to convince many in the field that EP's use of natural selection makes is sufficient as an explanation. Furthermore, the biological sciences have moved on since the 1990s and biologists

[7] Interestingly, although Baron-Cohen was once somewhat of a poster child for Evolutionary Psychology, Baron-Cohen himself has moved on to other aspects of autism research and theory that, while perhaps in some ways similar in form to EP-styled explanation, does not explicitly cast itself as EP.

[8] "Mindblindness" is a term that Baron-Cohen (1995) in fact coined for autism spectrum disorder. It is notable that social cognition is only one aspect of what is often compromised in ASD.

routinely speak of non-genetic and epigenetic channels of inheritance that are clearly relevant to functional organization (Jablonka and Lamb 2005), and which are processes that should presumably be taken seriously in any evolutionary account of human cognition.

3.3 Core Knowledge Theory: Massively Modular and Massively Innate

The so-called core knowledge explanation of human development postulates the existence of domain-specific ("core") competencies upon which later skills critically depend. Core knowledge theories have become a bit of a cottage industry in the past 20 years or so with core domains theorized to exist for number concepts, language concepts, physical concepts such as gravity and containment relations, folk biology, folk psychology, and most recently moral intuitions. Ironically, although EP-styled forms of explanation seem to be falling out of favour in psychology, core knowledge theories have become more common and influential. This is surprising because although core knowledge approaches are less explicitly "evolutionary", they are equally domain-specific, make similarly selectionist arguments—and cite the work of evolutionary psychologists for support. We surmise that the reason that core knowledge accounts seem to be thriving in our psychology when EP might be on the way out is the apparent empirical rigour of core knowledge accounts. However, as with Trevarthen, core knowledge accounts are explicitly innatist. And, as with Trevarthen, this aspect of their theorizing is quite out of step with recent work in the biological sciences (Lickliter and Honeycutt 2013).

However, the early social cognitive work of Baillargeon and colleagues has provoked quite a bit of recent discussion in psychology (for a review, see Baillargeon et al. 2010). Their account of false belief understanding has invigorated early social cognitive research since Onishi and Baillargeon's (2005) landmark study in which they concluded, using familiarization and looking time (so-called violation of expectation measures), that 15-month-old infants understand false beliefs. There has since been a study conducted using a modified version of the Baillargeon paradigm that fixed false belief understanding at 10 months (Luo 2011). The innovation—and limitation—of core knowledge theories are that they rely on presenting conceptually relevant stimuli (e.g. a situation depicting false belief) and then determine if infants look surprised when an actor violates this expectation. In this sense, they rely on a perceptual rather than a conversational paradigm where children can answer questions about false beliefs and make correct predictions (or not) about a protagonist.

Baillargeon et al. (2010: 111) "assume that infants are born with a psychological reasoning system that provides them with a skeletal causal framework for interpreting the actions of others". Like Trevarthen, they implicitly rely on a Chomskyan poverty of the stimulus form of argumentation, which draws its force in part from the claim that the skills in question develop very early and,

consequently, suggest unlearned and domain-specific knowledge. However, the fact that something appears early does not, of course, mean it cannot in some sense be learned. But, in either case, this simple dichotomization of learning and acquired is also a form of nature–nurture reasoning that we doubt anyone would want to defend in public (Bateson and Mameli 2007). When one adds to this the fact that their conclusions only follow if core knowledge researchers are actually investigating an earlier developing form of a logically identical behaviour—and that there are no clear mapping rules from their perceptually based studies to the conceptually based ones performed in other paradigms—it gets even more puzzling. Explaining the false belief of a protagonist provides reasonable logical grounds for attributing at least a rudimentary, but possibly not adult-like, understanding of false beliefs to a child. However, it is not at all clear that the looking longer at a violated false belief situation entails anything of the sort (Müller and Racine 2010). Moreover, given that the associations assessed in this paradigm could well be learned, there is little empirical force to this argument anyway.

Although some core knowledge researchers make some appeals to evolutionary logic, for example Wynn (1998) pointing out that counting skills might have advantaged our hominid ancestors and adducing some comparative "counting" studies in other model species, Baillargeon and colleagues make no typical evolutionary arguments whatsoever. To the extent that they do so, it is through association with the work of evolutionary psychologists (Barrett et al. 2013). Therefore, we are left with the claim that infant social cognitive skills, which we would assume to include joint attention, must have evolved because they develop early and cross-culturally. This inference rests, however, on a notion of "innate" that Bateson and Mameli (2007) have taken pains to show is problematic: early and robust development, even when it involves little apparent learning, need not imply innateness in the sense of an inner biological constraint.

3.4 The Shared Intentionality Hypothesis: Explicitly Adaptationist and Implicitly Innatist

By contrast, Tomasello and colleagues rely heavily upon comparative research in human infants and chimpanzees for their evolutionary account of joint attention, and make the case for the existence of many shared capacities between the two lineages as well as important differences that they contend account for the unique psychological capacities of humans. Tomasello et al. argue that apes and humans share many social cognitive capacities, such as being able to gaze follow, knowing what an opponent in a competitive competition can and cannot see, and grant that captive chimpanzees can come to point for their caretakers to food that they want. Thus, they acknowledge that chimpanzees, like humans, understand psychological states, however, they do not understand as many states as humans, nor do they understand them to the same extent that humans do (Tomasello 2008; Tomasello et al. 2005, 2007). The key difference between humans and apes, they contend,

lies in an adaptation they call shared intentionality. Shared intentionality is a prosocial motivation to share attention with others, which they claim transforms capacities to share attention, exhibited both by humans and chimpanzees, into those psychological capacities unique to humans. Thus, shared intentionality "is a big part of what makes humans unique in the animal kingdom, serving as a psychological foundation for all things cultural" and "skills and motivations for shared intentionality are…direct expressions of the biological adaptation that enables children to participate in the cultural practices around them" (Tomasello and Carpenter 2007: 124).

The claim that a phenotype is a "direct expression" of an adaptation seems to militate against the sort of moderation and dismissal of unqualified innatism that Tomasello and colleagues argue against in other places. They claim, for example, (e.g. Tomasello et al. 2005: 688) that "to understand the origins of a human cognitive skill we must go beyond simply labelling it as 'innate'. Indeed, although we concur that understanding actions as goal directed is a biological adaptation, this says nothing about the ontogenetic process". We obviously agree. However, in the same way that "innate" says nothing about how a phenotype develops, it also tells us nothing about how the phenotype evolved.[9] The underlying problem is that despite appearances and occasional appeals to the contrary, innatist claims are actually not, as they stand, evolutionary claims. If innatist claims are understood in the typical selectionist and adaptationist manner in which they frequently are, minimally, some plausible account of differential reproductive success is required. Here, Tomasello (2008) does quite well with an account of how our evolutionary ancestors might have benefitted from their shared intentionality, but of course, these sorts of "just so stories" as Gould and Lewontin (1979) (and Kipling) called them, have their own issues, and what seems to pass evolutionary muster in psychology would probably not in the biological sciences (Racine 2013).

4 Beyond Innate and Adaptation: The Reintegration of Evolution and Development

We have argued that "innate" is simply a vacuous concept especially when used as an explanatory device (e.g. Bateson and Mameli 2007; Racine 2013), and is essentially an admission that one does not understand how a given phenotype develops or has evolved, and is at best a promissory note. Unqualified adaptationist claims run similar risks for reasons that were well documented by Gould and Lewontin; it is simply tricky to reverse engineer an adaptation from a presently adaptive behaviour. As Gould and Lewontin (1979) note, previous adaptations can be co-opted to take on new functions and presently adaptive behaviours

[9] Of course, this should not be taken to mean that we dispute that there are evolved adaptations, or that we believe that processes of adaptation are somehow unimportant in natural selection.

can have no corresponding selective regime. Gene frequencies also change as a result of a variety of well-understood processes such as genetic drift, gene migration, and assortative mating, which do in some cases have clear phenotypic effects. This is not even to consider newer work in the biological sciences that has begun to reintegrate evolution and development (e.g. evolutionary developmental biology or "evo-devo"). However, the neo-Darwinian integration of Darwinian natural selection and Mendelian genetics, in which many innatist claims are purportedly grounded, has been very successful in the biological sciences. Progress though often comes with a price.

The price in this case was that, in order for the population genetic model that underwrites the modern synthesis of Darwin and Mendel to work, reliable developmental processes were simply presupposed (Lickliter and Honeycutt 2013). This reflected the separation of the science of development, or embryology as it was then called, from the study of heredity that was occurring at the same time. Although there are many interesting details to this that we do not have the space to discuss, what is most relevant is that neo-Darwinian thinking, despite Darwin's own interest in development, came to exclude development from evolution. This is reflected in Mayr's (1961) separation of ultimate versus proximate causation, and Tinbergen's (1963) explicit addition of ontogeny to Huxley's three problems of biology, which added a distinction between phylogeny and ontogeny to the intellectual landscape. Although these are reasonable assumptions to make if one's goal is to characterize changes in allele frequencies in a population, the selectionist and adaptationist mode of explanation has been criticized for its lack of use in explaining the developmental emergence of phenotype, and for its potential genetic determinism and reductionism (Lickliter and Honeycutt 2013; Racine 2012a; Racine et al. 2012; Wereha and Racine 2012).

Although epigenetics has become an important topic in the biological and psychological sciences in the past decade, and psychologists such as Gottlieb (2002) have used the term "epigenesis" in a related but historical sense to indicate opposition to preformationist views of phenotype, and to emphasize the importance of developmental processes in the construction of phenotype, most of the shifts in thinking that characterize the current evolutionary theoretical landscape have barely made their way into psychology at all. It is with more than a touch of irony, then, that in recent years adaptationist thinking has become increasingly common in development and comparative psychology, and joint attention theory in particular. All three forms of evolutionary theorizing reviewed earlier are implicitly or explicitly predicated on a neo-Darwinian view whose limitations are more fully recognized outside of the discipline.

In essence, it is not just EP, but adaptationist approaches in general that focus on the "design features" of a given phenotype by reasoning about the particular problem the putative adaptation might have solved in an ancestral environment. Although the empirical foundation for much of the core knowledge and shared intentionality work is creative and empirically sound, if developmentalists who explicitly—or even implicitly—draw on evolutionary theory wish to avoid Gould and Lewontin's (1979) "just so story" characterization of adaptationist

explanation, they need to be avoid unqualified uses of concepts like “innate”, “adaptation”, “adaptive”, “instinctual”, or “designed”, and the adaptationist forms of thinking with which they are tied up.

The first place to start is realizing that calling a phenotype the product of an adaptation when attempting to explain developmental processes is to not only give no explanation, but is to use neo-Darwinian ideas for a purpose for which they were not constructed. Second, at least some of the theoretical tools found in developmental systems approaches that consider the tight relation between developmental and evolutionary processes are already in place, and it is clear that the pioneering work of Oyama, Gottlieb, Lickliter, and others is exerting an influence in psychology. Ironically though, developmental systems approaches are often understood to be ways of taking into account the complexity of developmental processes without appreciating that systems views have an implicit evolutionary underbelly. The negative evolutionary critique of DST is one that is critical of EP, core knowledge, and shared intentionality’s predetermined epigenetic notions of adaptation (i.e. adaptations that arise from developmental information in sets of genes). A developmental systems perspective calls for more complete explanations and rigorous science. What this means in clear practical terms is conducting careful work that involves many levels of explanation—coupled with the realization that this is part of evolutionary explanation (Lickliter and Honeycutt 2013). In terms of social communication and particularly “high-level” abilities like language, we need to study more basic, but logically related, abilities like joint attention because such work is intrinsic to constructing plausible evolutionary explanations of the more complex abilities (Whiten and Erdal 2012).

In the rejection of the familiar nature–nurture and innate–acquired dichotomies which many psychologists, including those of EP, core knowledge, and shared intentionality stripes claim to agree on, the evolutionary and developmental landscape is changing in such a way that development–evolution becomes another problematic dichotomy that is impeding scientific progress. If so, the current generation of researchers and theoreticians will be the beneficiaries.

References

Akhtar N, Gernsbacher MA (2007) Joint attention and early vocabulary: a critical look. Lang Linguist Compass 1:195–207

Andrews K (2012) Do apes read minds? Toward a new folk psychology. MIT Press, Cambridge

Baillargeon R, Scott RM, He Z (2010) False-belief understanding in infants. Trends Cogn Sci 14:110–118

Bakeman R, Adamson LB (1984) Coordinating attention to people and objects in mother-infant and peer-infant interaction. Child Dev 55:1278–1289

Baron-Cohen S (1995) Mindblindness: an essay on autism and theory of mind. Cambridge University Press, Cambridge

Barrett HC, Broesch T, Scott R, He Z, Baillargeon R, Wu D, Bolz M, Henrich J, Setoh P, Wang J, Laurence S (2013) Early false-belief understanding in traditional non-western societies. Proc R Soc B Biol Sci 280:1–6

Bates E, Camaioni L, Volterra V (1975) The acquisition of performatives prior to speech. Merrill-Palmer Q 21:205–226
Bateson P, Mameli M (2007) The innate and the acquired: useful clusters or a residual distinction from folk biology? Dev Psychobiol 49:818–831
Boyd R, Richerson PJ, Henrich J (2011) The cultural niche: why social learning is essential for human adaptation. Proc Natl Acad Sci USA 108:10918–10925
Brauer J, Call J, Tomasello M (2005) All great ape species follow gaze to distant locations and around barriers. J Comp Psychol 119:145–154
Brinck I, Liljenfors R (2013) The developmental origin of metacognition. Infant Child Dev 22:85–101
Brown S (1999) The "musilanguage" model of music evolution. In: Wallin NL, Merker B, Brown S (eds) The origins of music. MIT Press, Cambridge, pp 271–301
Bruner JS (1983) Child's talk: learning to use language. Norton, New York
Butterworth G (1998) What is special about pointing in babies? In: Simion F, Butterworth G (eds) The development of sensory, motor and cognitive capacities in early infancy: from perception to cognition. Psychology Press, Hove, pp 29–40
Butterworth G (2003) Pointing is the royal road to language for babies. In: Kita S (ed) Pointing: where language, culture, and cognition meet. Erlbaum, Mahwah, pp 9–33
Call J, Tomasello M (2008) Does the chimpanzee have a theory of mind? 30 years later. Trends Cogn Sci 12:187–192
Carruthers P (2009) How we know our own minds: the relationship between mindreading and metacognition. Behav Brain Sci 32:121–138
Charles EP, Rivera SM (2009) Object permanence and method of disappearance: looking measures further contradict reaching measures. Dev Sci 12:991–1006
Tooby J, Cosmides, L (1995) Foreword. In: Baron-Cohen S (ed) Mindblindness: an essay on autism and theory of mind. Cambridge University Press, Cambridge, pp xi–xviii
Darwin C (1879) The descent of man, and selection in relation to sex, 2nd edn. John Murray, London
Gallese V (2001) The shared manifold hypothesis: from mirror neurons to empathy. J Conscious Stud 8:33–50
Glock H-J (2013) Animal minds: philosophical and scientific aspects. In: Racine TP, Slaney KL (eds) A Wittgensteinian perspective on the role and use of conceptual analysis in psychology. Palgrave Macmillan, Basingstoke, pp 130–152
Gottlieb G (2002) Individual development and evolution: the genesis of novel behaviour. Erlbaum, Mahwah
Gould SJ, Lewontin RC (1979) The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc R Soc Lond B 205:581–598
Hauser MD, Chomsky N, Fitch WD (2002) The faculty of language: what is it, who has it, and how did it evolve? Science 298:1569–1579
Heyes C (2012) New thinking: the evolution of human cognition. Philos Trans R Soc B 367:2091–2096
Hutto DD, Myin E (2013) Radicalizing enactivism: basic minds without content. MIT Press, Cambridge
Jablonka E, Lamb MJ (2005) Evolution in four dimensions: genetic, epigenetic, behavioural and symbolic variation in the history of life. MIT Press, Cambridge
Jones SS (2009) The development of imitation in infancy. Philos Trans R Soc B 364:2325–2335
Leavens DA (2012) Pointing: contexts and instrumentality. In: Pika S, Liebal K (eds) Developments in primate gesture research. Benjamins, Amsterdam, pp 181–197
Leavens DA, Racine TP (2009) Joint attention in apes and humans: are humans unique? J Conscious Stud 16:240–267
Lickliter R, Honeycutt H (2013) A developmental evolutionary framework for psychology. Rev Gen Psychol 17:184–189
Luo Y (2011) Do 10-month-old infants understand others' false beliefs? Cognition 121:289–298

Mayr E (1961) Cause and effect in biology. Science 134:1501–1506
Morissette P, Ricard M, Décarie TG (1995) Joint visual attention and pointing in infancy: a longitudinal study of comprehension. Br J Dev Psychol 13:163–175
Müller U, Racine TP (2010) The development of representation and concepts. In: Overton WF, Lerner RM (eds) Handbook of lifespan human development, vol 1., Methods, biology, neuroscience and cognitive developmentWiley, Hoboken, pp 346–390
Onishi KH, Baillargeon R (2005) Do 15-month-old infants understand false beliefs? Sci 308:255–258
Oyama S (2000) The ontogeny of information: developmental systems and evolution, 2nd edn. Duke University Press, Durham
Penn DC, Holyoak KJ, Povinelli DJ (2008) Darwin's mistake: explaining the discontinuity between human and nonhuman minds. Behav Brain Sci 31:109–130
Pika S (2008) Evidence for intentional and referential communication in great apes? In: Zlatev J, Racine TP, Sinha C, Itkonen E (eds) The shared mind: perspectives on intersubjectivity. Benjamins, Amsterdam, pp 165–186
Pika S, Bugnyar T (2011) The use of referential gestures in ravens (Corvus corax) in the wild. Nature Commun 2:560
Pinker (1994) The language instinct: the new science of language and mind. Penguin, New York
Pinker S (2010) The cognitive niche: coevolution of intelligence, sociality and language. Proc Natl Acad Sci USA 107:8993–8999
Place UT (2000) The role of the hand in the evolution of language. Psycoloquy. http://www.cogsci.soton.ac.uk/psyc
Povinelli DJ, Davis DR (1994) Differences between chimpanzees (Pan troglodytes) and humans (Homo sapiens) in the resting state of the index finger: implications for pointing. J Comp Psychol 108:134–139
Povinelli DJ, Bering J, Giambrone S (2003) Chimpanzee 'pointing': another error of the argument by analogy? In: Kita S (ed) Pointing: where language culture and cognition meet. Erlbaum, Mahwah, pp 35–68
Racine TP (2012a) Cognitivism, adaptationism, and pointing. In: Pika S, Liebal K (eds) Developments in primate gesture research. Benjamins, Amsterdam, pp 163–178
Racine TP (2012b) Getting beyond rich and lean views of joint attention. In: Seemann A (ed) Joint attention: new developments in psychology, philosophy of mind, and social neuroscience. MIT Press, Cambridge, pp 21–42
Racine TP (2013) How useful are the concepts 'innate' and 'adaptation' for explaining human development? Hum Dev 56:141–146
Racine TP, Carpendale JIM (2007) The role of shared practice in joint attention. Br J Dev Psychol 25:3–25
Racine TP, Wereha TJ, Leavens DA (2012) To what extent nonhuman primates are intersubjective and why. In: Foolen A, Lüdtke U, Racine TP, Zlatev J (eds) Moving ourselves, moving others: motion and emotion in intersubjectivity, consciousness and language. Benjamins, Amsterdam, pp 221–242
Slaney KL, Racine TP (2011) On the ambiguity of concept use in psychology: Is the concept 'concept' a useful concept? J Theor Philos Psychol 31:73–89
Tallis RF (2011) Aping mankind: neuromania, Darwinitis, and the misrepresentation of humanity. Acumen, Durham
Tinbergen N (1963) On aims and methods of ethology. Zeitschirift für Tierpsychologie 20:410–413
Tomasello M (1995) Joint attention as social cognition. In: Moore C, Dunham P (eds) Joint attention: its origins and role in development. Erlbaum, Hillsdale, pp 103–130
Tomasello M (2006) Why don't apes point? In: Enfield NJ, Levinson SC (eds) Roots of human sociality: culture, cognition and interaction. Berg, Oxford, pp 506–524
Tomasello M (2008) Origins of human communication. MIT Press, Cambridge
Tomasello M (2014) A natural history of human thinking. Harvard University Press, Cambridge, MA
Tomasello M, Carpenter M (2007) Shared intentionality. Dev Sci 10:121–125

Tomasello M, Carpenter M, Call J, Behne T, Moll H (2005) Understanding and sharing intentions: the origins of cultural cognition. Behav Brain Sci 28:675–735

Tomasello M, Carpenter M, Liszkowski U (2007) A new look at infant pointing. Child Dev 78:705–722

Tooby J, Cosmides L (1992) The psychological foundations of culture. In: Barkow J, Cosmides L, Tooby J (eds) The adapted mind: evolutionary psychology and the generation of culture. Oxford University Press, New York, pp 19–136

Trevarthen C (1979) Communication and cooperation in early infancy: a description of primary intersubjectivity. In: Bullowa M (ed) Before speech: the beginning of interpersonal communication. Cambridge University Press, Cambridge, pp 321–347

Trevarthen C (2011) What is it like to be a person who knows nothing? Defining the active intersubjective mind of a newborn human being. Infant Child Dev 20:119–135

Trevarthen C, Hubley P (1978) Secondary intersubjectivity: confidence, confiding, and acts of meaning in the first year. In: Lock A (ed) Action, gesture, and symbol: the emergence of language. Academic Press, London, pp 183–229

Wereha TJ, Racine TP (2012) Evolution, development, and human social cognition. Rev Philos Psychol 3:559–579

Whiten A, Erdal D (2012) The human socio-cognitive niche and its evolutionary origins. Philos Trans R Soc B 367:2119–2129

Wilkins D (2003) Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). In: Kita S (ed) Pointing: where language, culture, and cognition meet. Erlbaum, Mahwah, pp 171–216

Wynn K (1998) An evolved capacity for number. In: Cummins D, Allen C (eds) The evolution of mind. Oxford University Press, New York, pp 107–126