Agency and Joint Attention

Edited by Janet Metcalfe
and Herbert S. Terrace

OXFORD UNIVERSITY PRESS 2013
How Joint Is the Joint Attention of Apes and Human Infants?
Malinda Carpenter and Josep Call

Recently there has been much debate about whether chimpanzees and other apes engage in joint attention. Some researchers (e.g., Leavens & Racine, 2009; Tanner & Byrne, 2010) argue that they do, whereas others (e.g., Tomasello, Carpenter, Call, Behne, & Moll, 2005) argue that they do not. A similar debate is emerging in developmental psychology, as some researchers argue that joint attention emerges ontogenetically around 9 to 12 months of age (e.g., Tomasello, 1995), whereas others argue that it emerges much earlier (e.g., Grossmann & Johnson, 2010; Striano & Bertin, 2005). However, different participants in these debates often use different definitions of joint attention, a problem that likely is responsible for much of the disagreement in both areas. Thus, in the hopes of eventually clearing up some of these disagreements, in this chapter we first argue for the adoption of a relatively conservative definition of joint attention—one that requires the coordination of attention in joint attention to be truly joint. Then we examine the empirical evidence relevant for determining whether apes—and when human infants—engage in joint attention under this definition. We conclude that whereas human infants engage in truly joint joint attention by 1 year of age, there is so far little if any convincing evidence that chimpanzees and other apes do. We end by discussing possible reasons for differences between humans and apes in joint attention and other related abilities such as joint action.
The classic definition of joint attention involves a triadic interaction in which two individuals coordinate attention to an object of mutual interest (e.g., Bakeman & Adamson, 1984). Across time, this definition has been pulled in different directions by different researchers. Some focus more on the "triadness" and the fact that the two individuals are looking at the same thing, whereas others focus more on the coordination aspect of it and the active sharing of attention. For example, in the former group, Leopold and Racine (2009, p. 241) define joint attention as "the intentional co-orientation of two or more organisms to the same locus," with "at least one of the organisms" being something intentionally so as to end up focusing on the same thing as the other. They take a variety of triadic behaviors as evidence of joint attention; for example, gaze following, gesturing about objects (including requests), and social referencing. In contrast, the latter group of researchers (including the authors) argue that attending to the same thing that one's partner is attending to is not enough for joint attention: In addition, it is crucial that both partners know together that they are attending to the same thing (see, e.g., Tomasello, 1999). In this view, both partners are (at least eventually) equally involved, actively sharing attention about the thing. This is what makes joint attention joint, rather than just parallel attention to the same object.

To help see the difference between these two approaches, imagine a gaze following situation in which one individual sees another turn to look at something, and then turns to look at it himself. Both individuals are now simultaneously attending to the same thing. But where is the jointness here? They may be looking at the same thing but they are not necessarily doing so together. Gaze following can be done in a unilateral, even exploitative manner, and the looker need not even know that the follower is present, much less sharing attention (note that some other triadic behaviors such as social referencing can work in this unilateral, exploitative way as well). Similarly, imagine a situation in which one individual uses a gesture like a reach or point to direct another individual's attention to an object. Again, they may end up looking at the same object simultaneously but where is the jointness here? The gesture could simply want the recipient to see the object individually, for example, so that the recipient gives it to him or is informed about it. Thus, many triadic behaviors can involve parallel, rather than joint or shared attention.

Let us be clear that we are not claiming that behaviors such as gaze following and attention-directing gestures are not joint attention, just that they are not necessarily joint attention. There are clearly cases in which these behaviors can involve participants knowing together that they are sharing attention. The challenge for researchers who claim that chimpanzees or very young infants engage in joint attention is thus to provide evidence of this knowing together—the third leg of the "joint attentional triangle" (see Fig. 4.1).

The most commonly used evidence of this knowing together or sharing of attention is gaze alternation between the object of interest and the eyes/face of the partner. However, gaze alternation alone is not enough to prove the existence of joint attention because there are many situations in which one might look back and forth between an object and a social partner without coordinating attention with that individual. Gaze alternation can be a sign of alternating or checking attention, rather than of coordinated attention (Tomasello, 1999).

If it is so difficult to know whether two individuals are engaging in joint attention, then how do we know that human infants (at any age) do it? One approach is to look for interactions in which the sole purpose is to share attention about objects or events: joint attentional engagement and declarative gestures (in other words, participants' production of joint attention behaviors). Another approach is to conduct experiments aimed at determining whether participants recognize whether they have shared attention with others about something (i.e., participants' comprehension of whether they have been in joint attentional engagement with someone). We turn to both of these types of evidence now, focusing (very briefly) on a few studies that have addressed this question directly.

Evidence for Joint Attentional Engagement, Declarative Gestures, and "Knowing Together" in Human Infants

By 9 months of age, most typically developing human infants have started participating in interactions thought to involve joint attentional engagement (e.g., Carpenter, Nagell, & Tomasello, 1998). The classical operational definition of (coordinated) joint engagement involves a look to a caregiver's face/eyes about an object or event in the infant's and caregiver's mutual focus of attention (i.e., an object that they both have been looking at, and will turn back to again afterward) (e.g., Bakeman & Adamson, 1984). However, we think it likely that this type of operational definition may sometimes result in false positives, and thus that coding schemes that capture the quality of the looks to the social partner, like the one Hobson and Hobson (2007) used with much older children to distinguish "sharing" from "checking" and "orienting" looks, would be more useful in identifying true joint attentional engagement. When infants look to the adult in these
interactions are true sharing looks (e.g., see Fig. 2.1b), we think previous researchers are justified in describing these interactions as a “meeting of minds” (e.g., Bruner, 1993).

Luckily, at around the same age as they begin initiating mutual gaze about objects, infants begin to produce so-called declarative gestures, and this makes the diagnosis of joint attention much more straightforward. Infants’ first declarative gestures are usually shows (i.e., holding up an object toward an adult’s face with a look and smile to the adult), then around age 12 months infants start pointing to more distal objects (again, with accompanying look and smile to the adult) to share attention and interest about the objects with the adult (e.g., Carpenter et al., 1998; for more on declarative pointing, see chapter 8).

The theory is that with both joint attention looks and declarative gestures, sharing attention with the other person is an end in itself (e.g., Gómez, Sarití, & Tamari, 1995). However, there are other plausible interpretations of these behaviors, so it is important to test empirically what infants are attempting to do when they perform them. One way to test this is to experimentally manipulate the adult’s reaction to infants’ looks or gestures, in order to rule out possible lower-level explanations. Most work in this area has been done on declarative pointing because it is so easy to elicit and code in infants. For example, following Perucchi and Camapלתi (1993), one way to know whether infants are pointing simply because they want to obtain the object they are pointing to, instead of to share attention, is to have the adult respond to the infants’ points with just a comment and smile ("Oh, that’s nice!"). If infants are satisfied with this reaction, then that is apparently all they wanted from her, whereas if they keep pointing, then they probably want to obtain the object. Using this method, Carpenter et al. (1998) found that declarative pointing emerged at 12 months, on average, in their longitudinal sample. Similarly, we have tested whether infants are pointing simply to gain rewarding positive emotions to the self (Bates et al., 1997; Moore & D’Entremont, 2001) or just so that the adult sees the object for herself, rather than to share attention about the object. We found that if an adult responds to 12-month-olds’ points by speaking enthusiastically with positive emotion only to infants (ignoring the object), infants are apparently unsatisfied and repeat their pointing. They show the same response if the adult just looks at the object silently. However, if the adult responds with joint attention (alternating gaze between the infant and the object while speaking with positive emotion), infants are satisfied with her reaction and do not repeat their point (Liszkowski, Carpenter, Henning, Sterlano, & Tomasello, 2004).

Further studies show that 12-month-olds repair adult misunderstandings about the precise referent of their declarative points (Liszkowski, Carpenter, & Tomasello, 2007b), so their points are clearly about something specific. Twelve-month-olds can even point declaratively about absent referents—objects that were previously present but that have now been taken away (Liszkowski, Carpenter, & Tomasello, 2007b)—showing that the sharing of attention and interest takes place on a mental level (and is not simply about having the adult physically turn and look at something). Finally, by 18 months of age, infants point selectively to referents that are relevant to their previously shared experiences with the particular person for whom they are pointing (Liebal, Carpenter, & Tomasello, 2010). Thus, 1-year-old infants show great flexibility and complexity in their declarative pointing, inviting others to share attention with them for specific, relevant, and even currently perceptually absent things.

But what evidence is there that infants at this age understand that they and the adult know together that they are sharing attention? The most direct evidence of this comes from a study by Moll, Richter, Carpenter, and Tomasello (2002). They had 14-month-olds share (i.e., interact excitedly about) three objects with an adult in sequence, one of them in a special way (they encountered it several times on the way to the testing room). When later the adult gestured excitedly toward the three objects together on a tray and ambiguously requested, "Wow, look, can you give it to me please?" infants gave her the object they had shared in a special way. To test the possibility that infants simply gave the adult that object because it was special for them, individually (not because it was the one they had shared together), Moll and colleagues included a control condition in which infants shared the three objects with the adult exactly as before, one in a special way, but then a different adult ambiguously requested one of the objects in exactly the same manner. If infants were just choosing the special object because it was most interesting to them, they should have chosen it in this condition too, but they did not—they chose that object at chance levels. To test the possibility that infants gave the adult the special object because it was the object that was apparently special for her, in another control condition infants watched as the adult experimented the objects individually (again, one in a special way) and then requested one of the objects from them ambiguously. Again, in this condition infants chose the special object at chance levels. In summary, infants responded not based on what they themselves knew about the objects, nor based on what the adult knew individually, but instead based on what they knew together with the adult. By 14 months, infants thus know what "we" know together (see also, e.g., Liebal, Behne, Carpenter, & Tomasello, 2009; Liebal et al., 2010; Saylor & Ganea, 2007; for further evidence of this).
How Joint Is the Joint Attention of Apes and Human Infants?

Younger Infants

There are now a number of studies that have looked for production and comprehension of joint attention in infants younger than 9 months of age. To take just two examples, Striano and Beren (2005) found evidence of gape alternation between objects and adults as young as 5 and 7 months of age, and Grossman and Johnson (2010) showed that 3-month-olds are sensitive to whether interactions are triadic or not (see chapter 13 for a more in-depth review). Of course, infants from very early on are able to follow other's gaze, at least to nearby objects (e.g., D'Entremont, Haines, & Mui, 1997). However, we know of no evidence that infants this young know together with others that they are sharing attention; so far the best evidence we have of this begins at 12-14 months (see the preceding section). Future research is clearly needed on this topic.

Chimpanzees and Other Apes

Chimpanzees follow others' gaze, and can do so in quite sophisticated ways (see Call & Santos, 2012; Call & Tomasello, 2008 for reviews). For example, they can follow gaze to locations behind themselves and behind barriers (e.g., Povinelli & Eddy, 1996; Tomasello, Call, & Hare, 1998), and they only fail to follow gaze through barriers that the looker can see through (Okamoto-Barth, Call, & Tomasello, 2007). They can follow both head and eye direction (Tomasello, Hare, Lehmann, & Call, 2007), and when they follow gaze and do not see anything interesting, they look back to double check the looker's line of regard (Briner, Call, & Tomasello, 2005; Call, Hare, & Tomasello, 1998). Thus, chimpanzees have quite highly developed skills of gazing following, and are apparently motivated to attend to the same thing to which others are attending.

Chimpanzees also direct others' attention to things to which they themselves are attending. They routinely gesture toward objects for human caregivers in order to get them to attend to the objects and act on them (see Call & Tomasello, 2007, for an overview of ape gestures). These attention-directing gestures are quite sophisticated too. They are intentional (e.g., Gómez, 1996; Leavens, Russell & Hopkins, 2005; Tomasello, Call, Nagell, Olguin, & Carpenter, 1994) and referential (e.g., Leavens, Hopkins, & Thomas, 2004; Menzel, 1999). In addition, chimpanzees often alternate gaze between the object and the recipient while communicating (Leavens et al., 2004) and they choose different gestures depending on the attentional state of the recipient, using visual gestures for recipients who are looking at them but auditory or tactile gestures more often for recipients who are not already looking at them (Tomasello et al., 1994, 1997). Chimpanzees are thus clearly both able and motivated to direct others' attention in sophisticated ways to objects they are attending to in their environment.

As we have already explained, however, gaze following and attention-directing gestures can involve individual, parallel attention to an object, rather than joint attention. We thus need to look for additional evidence, again, of (1) a motivation to share attention with others with no other more instrumental goals, and (2) both individuals knowing together that they are sharing attention. For the first point, researchers from several different labs have concluded that apes do not appear to share attention and interest with others in joint attentional engagement (e.g., Bard & Vauclair, 1984; Tomasello & Carpenter, 2003; Tomonaga et al., 2004) or by using declarative gestures such as looks and declarative points (e.g., Gómez et al., 1993; Tomasello & Carpenter, 1995; Tomonaga et al., 2004). There are a few papers that report evidence of joint attention and declarative gestures, however. Next we detail what that evidence consists of, and why we find it unconvincing.

With regard to joint attentional engagement, one of us has reported that chimpanzees and bonobos spend less time in joint attentional episodes than 18-month-old infants, but the apes were still credited with some joint attention in that study (Carpenter, Tomasello, & Savage-Rumbaugh, 1995). However, for practical reasons, in that study we simply coded looks to the experimenter's face and the objects present, and it was our impression at the time that the rather sterile nature of the coding—which did not consider why participants looked to the experimenter's face, just whether and when they did—almost certainly led to many false positives in terms of true joint attention with sharing looks. In an and Byrne (2010) report triadic social play in gorillas, with joint attention and invitations to share interests in and attention to objects. However, although these interactions are clearly very social and do involve objects, we see little indication in the video clips of these interactions (available on Joanne Tanne's web page: http://www.gorillasources.info/CH_TRIADIC_PLAY_INTRO.htm) that the gorillas are sharing attention to the objects as an end in itself, the way human infants do. Instead, the objects seem mainly to be an incidental part of a basically dyadic, rough-and-tumble social interaction (similar in some ways to Baez and colleagues, 1975, original characterization of the use of an object to get the partner's attention to the self). Gómez (2010) also reports many observations of triadic social play between a young gorilla and her human caregivers, but notes that there was one main difference in the gorilla's looks to the humans, compared with the looks of human infants to their caregivers: Whereas human infants often look and smile to their caregiver after something interesting happens (as if to comment on it, we would argue—see Carpenter & Liebal, 2011), this gorilla did not do this. Instead, her looks to the humans occurred in the midst of social interactions like taking turns throwing things at each other. Thus, the message conveyed by the looks of the gorillas in both these studies seems to be something like, "Come on, let's play!" or "Do it again!" instead of something like, "Wow, wasn't that [object/event] interesting?" a typical joint attention response.

With regard to declarative gestures, Carpenter et al. (1995) reported two instances of possible declarative gestures by a human-raised bonobo, but both were extremely questionable in terms of the ape's intentions (again the coding was done based on superficial behavior, e.g., stretching out the leg with an object on it in the direction of the experimenter, and the coder was not convinced that the ape really was attempting to
How Joint Is the Joint Attention of Apes and Human Infants?

previous studies (e.g., Leavens, 2003; Leavens & Rarke, 1999; Leavens, Rarke, & Hopkins, 2009) listed numerous reports of what they say appear to be declarative points by apes. For example, gestures used to inform a human about the location of a tool that is needed to get a reward are considered declarative, as are warning signals, and trained responses to requests for information (e.g., a language-trained chimpanzee touching her nose after being asked to "show me your nose") - all apparently simply because they are not imperatives. There is also a single report of a single wild bonobo pointing (Ved & Sabater-Pi, 1998), which several authors have taken as evidence of declarative pointing in wild apes (although Ved & Sabater-Pi themselves do not claim that it is declarative pointing). However, in these reports we see no evidence of declarative gestures in the sense that this term is used with human infants: gestures performed with the motivation to share attention with others simply for the sake of sharing attention, nothing more (in the Ved & Sabater-Pi observation it appears to us to be a point to inform/warn others about the presence of humans hiding in a bush). Finally, Lyn, Greenfield, Savage-Rumbaugh, Gilletsie-Lynch, and Hopkins (2011) report an example of an enculturated chimpanzee pointing to a plane for an experimenter. It is simply not clear from the written record whether this was to inform the experimenter of the plane or to share attention with him or her.

Greenfield and Savage-Rumbaugh (1991) and Lyn et al. (2011) have also suggested that enculturated apes use lexigrams declaratively. In their studies, chimpanzees and bonobos used lexigrams not only to request food and activities, but also to respond to questions and announce what they were going to do next. These last two behaviors were considered declarative because they provide information to the recipient and do not involve a request. However, these responses were often explicitly trained, not spontaneous, and anyway again this is not the typical use of the term declarative, as it is used with human infants (i.e., simply to share attention to objects). It is important to keep in mind the distinction between informative gestures/utterances, which can be achieved with parallel attention, and declarative gestures/utterances, which require joint attention. In fact, other research shows that enculturated and human-raised apes left to their own devices produce very few gestures or lexigram "utterances" that would be classified as declarative - in any use of the term (e.g., Gómez, 2002; Gómez et al., 1999; Hayes, 1991; Rivet, 2003; Tomasello et al., 2004; see chapter 1). Thus, overall, even when using a very lenient and overly inclusive definition of declarative gestures, they are rare at best in nonhuman apes.

With regard to evidence of "knowing together" or comprehension of what has been shared with others, few relevant studies have been conducted with apes. There is some evidence suggesting that apes know what others know, in the sense of what others have seen in the immediate past (e.g., Hare, Call, & Tomasello, 2001; Kaminak, Call, & Tomasello, 2008; see also chapters 3 and 4), and there is even some evidence suggesting that they know what they themselves know in this sense (Call, 2010; Call & Carpenter, 2001). However, we are unaware of any studies that show that apes know what they know together with others. Future studies similar to those conducted with human infants are needed to investigate this (see, e.g., the Moll et al., 2008 study discussed previously).

SUMMARY

Whereas there is plenty of evidence (from both observational and experimental studies) that 1-year-old human infants engage in truly joint joint attention, there is little if any convincing evidence that chimpanzees and other apes do. Unlike humans, apes do not appear to participate in interactions involving joint attentional engagement, they do not produce declarative gestures serving to share attention with others, there is no evidence of comprehension of having been in joint attentional engagement (although more work needs to be done to address this latter point in particular). When they follow others' gaze, there is no indication that they know together or actively share the fact with the other individual that they are looking at the same thing (in contrast to human 1-year-olds, who do sometimes show this evidence, for example when they follow an adult's gaze and then point to the target object themselves with a look and smile to the adult, as if to comment on it or confirm with her that they have seen it; Brooks & Meszoff, 2003; Carpenter et al., 1998). Likewise, when apes gesture for others, there is no unequivocal evidence that they do so with the sole (and spontaneous) goal of sharing attention and interest with others about something, or with any consideration of what they have shared with others in the past. Even those researchers who claim to have found evidence of "triadic" interactions or declarative gestures admit that they are rare (e.g., Leavens & Rarke, 2003; Lyn et al., 2011; Tanne & Byrne, 2010), whereas in infants much more convincing behaviors are extremely frequent. Still, much work has been done on joint attention in apes with the more conservative definition of joint attention we advocate in mind, so the challenge to ape joint attention researchers is to provide this evidence if it can be found.

Why Don't Nonhuman Apes Engage in Joint Attention?

If apes do not engage in joint attention, why not? They have most of the prerequisites needed, both in terms of social-cognitive understanding and the physical behaviors involved: They are sensitive to what others see, know, and intend (e.g., Call & Tomasello, 2008), and they communicate with superficially identical gestures in other, imperative contexts (e.g., Call, 2009). Gómez et al. (1993) and Leavens et al. (2009) have proposed that nonenculturated apes may not point declaratively because of the impoverished relationships they have with their human caregivers, but this cannot be the full story because mother-reared apes (even those in the wild) do not point declaratively for their conspecifics either. Besides, the idea that only a proper human upbringing will enable apes to develop declarative pointing is not supported by the data (in that enculturated apes still
do not spontaneously communicate declaratively to share attention with others as an end in itself—and also clearly begs the question of why only humans provide this type of environment.

We think the answer to this question is that only humans have evolved the basic motivation to share psychological states with others. We think that apes not engaging in joint attention is part of a broader pattern involving also not participating in joint action (i.e., collaboration; Tomasello & Carpenter, 2007; Tomasello et al., 2005) and certain types of social imitation (Carpenter & Call, 2009). What all of these abilities have in common is the motivation to share or align psychological states with others. The parallel to joint action are particularly strong, and thus similar sorts of challenges apply for researchers claiming to have found joint action in apes (e.g., Boesch, 1985; Pika & Zuberbühler, 2003): Is there evidence of a shared goal—that apes know together with the other individual(s) that they are participating in a joint action, rather than simply engaging in parallel activity or using the other as a social tool? We are not erecting an insurmountable wall: These are questions that can be brought under empirical scrutiny, as the research done on human infants has shown. Here we simply hope to highlight the most important aspect of joint attention—its jointness—and recommend that it be an integral part of both its theoretical and operational definitions. This will help us determine whether apes and very young human infants engage in joint attention.

NOTE

1. Note that although Bates, Camarero, and Volterra (1975) coined the term proto-declarative gestures, the term is typically now used in a different, social-cognitively richer way than in their original account: as gestures used to direct others’ attention to objects for the purpose of sharing attention and interest in them.

REFERENCES

How Joint Is the Joint Attention of Ages and Human Infants?


