Do you see what I see? The neural bases of joint attention

Elizabeth Redcay & Rebecca Saxe

Department of Brain & Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA

In the simplest social interactions, two people look at each other. Gazing face-to-face is infants' first social experience with their parents, for example. Most social interactions between humans, however, don't stop with mutual gaze. Instead, people quickly switch to sharing attention on an object or topic of communication. I point out the flowers, you mention your sister's graduation, we both smile at the thought. In each of these cases, our social interaction has at least three elements: two people, coordinating attention (and perception, and emotional reactions) on a third element. These core elements of social interaction are called "joint attention" (or sometimes "triadic attention," to make the distinction from simple mutual gaze).

These triadic interactions provide an infant with a platform by which she can learn about her world. Similarly, they provide the caretaker with a platform by which she can guide the infants learning. Joint attention continues to be an important tool throughout life; however, it is particularly critical in infancy and early childhood when social interaction is the primary means of learning. Before an infant has acquired abstract symbols (words) for objects, her medium of communication (and learning) about her world with others is constrained to points,

gaze shifts, and head turns. The simple ability and motivation to share attention with another person on an object is highly correlated with a host of cognitive and social advances later in life. For example, joint attention ability is correlated with later language ability (Baldwin, 1991; Morales, Mundy, & Rojas, 1998; Mundy & Newell, 2007) emotion regulation (Morales, Mundy, Crowson, Neal, & Delgado, 2005), social competence (Vaughan Van Hecke, et al., 2007), and theory of mind (or reasoning about other's mental states) (Nelson, Adamson, & Bakeman, 2008).

Neuroscientists have only recently begun to investigate the neural mechanisms of joint attention. New non-invasive neuroimaging techniques now let scientists peer directly into the brains of healthy human adults, children and even infants, while they engage in simple social interactions. Neuroscience studies can inform what brain systems mediate a behavior. This neural information can often provide greater sensitivity than the study of behavior alone for several reasons. First, a behavior may be the same but the mechanism that produces it differs. For example, if I view a person look to my left or see an arrow point to my left, I will likely shift my own gaze to the left; however, orienting to a person's gaze relies on overlapping but distinct brain regions (and thus mechanisms) than does orienting to a non-social directional cue (Engell, et al., 2010). Second, neural sensitivity can serve as evidence for stimulus discrimination if an overt behavioral response is not reliably measurable. This is particularly true when studying infants and young children who may have more sophisticated abilities that are belied by their poor motor and attentional control. Finally, the identification of brain regions that

underlie a given behavior in typical, healthy individuals provides a baseline by which atypical populations can be compared.

Almost all neuroscience studies focus on a simple example of joint attention (although as we shall see, this example is complicated enough): person A sees person B looking at object C, and so person A shifts their visual attention to object C. Furthermore, neuroscience studies are largely restricted to the perspective of person A – the "responder". Below we will describe the few studies that have looked at the neural correlates for the "initiator" of joint attention. However, the bulk of this chapter will follow the literature, and try to characterize the cognitive and neural mechanisms that allow people to detect a bid for joint attention, and respond appropriately.

From the responder's perspective, engaging in joint attention requires at least four steps. First, the responder must attend to, or at least be aware of, the initiator as a potential social partner. Second, the responder must detect a shift in the attention of the initiator towards an object, for example, by perceiving the initiator's gaze shift, head turn and/or point. Third, the responder must shift attention to the object. And finally fourth, the responder must monitor the ongoing relationship between the initiator's attention, their own attention and the object, to ensure that joint attention is successfully achieved and maintained. While the first, second, and third components are relatively easy to isolate and examine in typical neuroimaging experiments, the fourth provides technical and methodological challenges. However, in many ways the fourth is the essence of joint attention. Joint

attention can only fully occur if two people actively coordinate attention and are aware of each other's attention on the same thing. In a typical functional magnetic resonance imaging (fMRI) experiment, participants lie flat on their back in a magnetic tube (the MRI bore). Activation is detected by looking for changes in MRI signal in small regions of the brain (about 45 mm^3); a resolution that requires participants stay motionless. Acquisition of these data produces a very loud repetitive and high-frequency sound. Let's remember now the goal of our study: to understand the neural bases underlying active coordination of attention with another person in the context of a social interaction. The challenges of creating a natural social interaction while in the scanner are daunting. As we will describe below, this challenge is only beginning to be overcome.

In the first section of this chapter, we describe what is known about the neural mechanisms of each of these four steps of joint attention. In the second section, we discuss how and whether neural evidence can help to address questions about the development of joint attention in human infancy.

Neural mechanisms of joint attention in adulthood

1. *Attending to a potential social partner*

In adults, attending to the presence of a social partner appears to depend on the dorsal medial prefrontal cortex (dMPFC). The dMPFC is located on the medial wall of the frontal lobe, anterior and superior to the cingulate gyrus; or about half an inch behind the middle of the forehead. Activity has been observed in the dMPFC

during a whole range of different tasks that require participants to think about, or interact with, another person.

Activity in the dMPFC is critical to a social interaction in two ways. First, it appears to reflect the sense that another person is *present* as a social partner and second, it supports reasoning about the person's psychological or emotional traits (Harris, Todorov, & Fiske, 2005; Mitchell, Heatherton, & Macrae, 2002; Mitchell, Neil Macrae, & Banaji, 2005), which may be necessary in a social interaction. DMPFC activity is specifically enhanced when the potential social partner deliberately draws the subject's attention to himself. A bid for attention can occur via direct mutual gaze (to catch your attention, I look you in the eye), or via calling the target's name. Both of these social cues elicit enhanced activity in the dMPFC (Kampe, Frith, & Frith, 2003). A particularly strong social cue occurs if the person walks towards the subject, looks towards them, and then makes a social gesture (e.g. smile, raised eyebrows); these cues elicit dMPFC activity (Schilbach, et al., 2006), but see (Pelphrey, Viola, & McCarthy, 2004). The duration of mutual gaze is also correlated with higher dMPFC activity (Kuzmanovic, et al., 2009).

In sum, in a joint attention episode, dMPFC activity may initially support attending to a potential social partner, especially if the social partner is deliberately eliciting that attention, signaling the beginning of an interaction.

2.   *Detecting a shift in attention*

Once the responder is attending to the initiator, the second step of a joint attention interaction occurs when the initiator shifts attention to another object.

The responder must detect and accurately represent that shift. To accurately do this, the observer must understand that the attentional shift was intentional. Considerable neuroscientific evidence, from both animal and human models, implicates the right posterior superior temporal sulcus (pSTS) in these calculations. The superior temporal sulcus runs the length of the temporal lobe; the region implicated in perceived shifts of attention, at least in humans, is near the posterior end, above the right ear.

In macaque monkeys, neurons in the pSTS respond to specific orientations of the head and eyes (Perrett, Hietanen, Oram, & Benson, 1992). Some of these neurons appear to literally code the observed direction of an actor's attention. For example, an individual neuron might prefer (i.e. discharge more action potentials when viewing an image of) a face in which the head is pointed toward location A but not B. If the head is inverted the same cell will fire only if the head is facing towards A (even though the direction of head orientation will be different) (Hasselmo, Rolls, Baylis, & Nalwa, 1989). Similarly, another neuron might prefer left-ward gaze (as compared to right-ward), regardless of the orientation of the head; however, if the eyes are occluded, the same neuron prefers a left-facing head to a straight or right-facing head, regardless of overall body orientation; and if the head is occluded, then the same neuron prefers a left-ward pointing body (Jellema, Baker, Wicker, & Perrett, 2000). These neurons thus appear to implement a very abstract code, representing the direction of another person's attention based on the best available evidence (eyes>head>body orientation). In humans, fMRI studies have similarly reported right pSTS activation when people observe (and attend to) another

person's gaze shifts (Hooker, et al., 2003; Kingstone, Tipper, Ristic, & Ngan, 2004; Materna, Dicke, & Thier, 2008; Pelphrey, Singerman, Allison, & McCarthy, 2003). Distinct regions of the posterior superior temporal sulcus (pSTS) appear to respond to motion in the eyes, mouth, and hands (Puce, Allison, Bentin, Gore, & McCarthy, 1998).

Importantly, the STS does not just respond to moving eyes or bodies, but appears to integrate the eye movement with the surrounding context in order to interpret the person's action. People don't just move their eyes; they look at specific objects. For example, in one experiment, participants saw an animated face that occasionally gazed towards one of the four corners of the screen. Immediately before the gaze shift, a flashing checkerboard would appear either in the target-of-gaze position (congruent) or somewhere else (incongruent). Both conditions elicited activation in the pSTS; however, the incongruent condition engaged the pSTS to a greater extent (Pelphrey, et al., 2003). That is, when gaze shifts are incongruous with the visible context, pSTS activity is enhanced (as if the pSTS is "working harder" to interpret the unexpected action). This incongruency effect has been shown with other types of actions, such as reaching and walking (Brass, Schmitt, Spengler, & Gergely, 2007; Pelphrey, Morris, & McCarthy, 2004; Saxe, Xiao, Kovacs, Perrett, & Kanwisher, 2004).

There is also some evidence that activity in the pSTS is enhanced when an action (such as gaze shift or point) reflects a communicative intention. In one study, activity in the pSTS was higher when an animated character with averted gaze

shifted his gaze towards the participant to make eye contact, compared to when he shifted his gaze further away from the participant (Pelphrey, Viola, et al., 2004). Thus, the pSTS may play a role in interpreting actions within a communicative context.

In sum, pSTS appears to be involved in perceiving and interpreting other people's biological actions, including particularly gaze and head orientation shifts. The pSTS allows the responder to detect and interpret the initiator's movements as evidence of a shift of her attention, providing the invitation for the responder to follow. The pSTS is recruited more if this biological action requires that the observer think about the intention behind the action.

3.   *Shifting one's own attention*

Another person's gaze shift is a powerful cue that elicits a shift in the observer's attention. This responsive shift of attention can occur both reflexively and automatically, and under deliberate control. To illustrate the difference: if you are focused on reading a chapter at your desk but colleagues are talking loudly outside your door, attention will be automatically oriented to your colleagues (i.e. exogenously). This is distinct from voluntary (endogenous) control of attention orienting such as subsequently choosing to focus back on your chapter. Exogenous and endogenous attention processes rely on distinct neural mechanisms (Corbetta & Shulman, 2002; Rosen, et al., 1999). Responding to joint attention involves both.

Orienting to another person's gaze or point has been shown to be a reflexive, automatic process, engaging the exogenous attention network. This has been tested

a number of times through variations of a simple cueing paradigm (Posner, 1980). In this paradigm, participants are told to push a button as soon as they see a target on the left or right side of the screen. Before the target appears, a face in the center of the screen will shift gaze to the right or left of the screen but the shift does not predict the location of the target. Even though the subjects are aware that the gaze cue is not informative, the subject will take longer to detect a cue if it appears on the side that was not cued by the gaze shift (Kingstone, et al., 2004; Stevens, West, Al-Aidroos, Weger, & Pratt, 2008). Exogenous control relies on a ventral frontoparietal network of brain regions, including the right inferior and middle frontal gyrus (R IFG and R MFG) and a region within the temporoparietal junction (TPJ) (Corbetta & Shulman, 2002). Thus, these regions play a role in automatic shifting of attention during perception of a gaze or point cue.

Although viewing another person's gaze automatically orients attention to where that person is looking, joint attention also requires controlled & voluntary shifts of attention (endogenous attention). This endogenous attention recruits a network of regions within the dorsolateral frontal parietal system. Two of these regions include the frontal eye fields (FEF) and intraparietal sulcus (IPS) (Corbetta & Shulman, 2002). This system is recruited during eye movements, as well as covert shifts of attention (Corbetta, et al., 1998; Nobre, Gitelman, Dias, & Mesulam, 2000). The FEF region plays a role in motor control and planning of eye movements while the intraparietal sulcus (IPS) plays a role in deployment and maintenance of spatial attention (Hopfinger, Buonocore, & Mangun, 2000).

Thus, both the exogenous (or reflexive) and endogenous (or controlled) attention systems rely on regions within frontal and parietal cortex which are distinct from those involved in other components of joint attention. Depending on the context, either system may be recruited to make a shift of attention (reflexive or controlled) during a joint attention interaction.

4. Monitoring the relationship between one's own attention, another's attention and an object (i.e. triadic attention).

After the responder shifts attention to an object, she must be able to detect and monitor that the initiator is also sharing attention with her. Joint, or triadic, attention only lasts as long as both people know that they are each attending to the object and each other. This knowledge or sense that you are sharing attention with another person *intentionally* is the key component in joint attention. This shared experience is what allows joint attention to be truly communicative and such a successful platform for learning.

Joint attention can be either concrete, such as physically looking at the same object in their shared environment, or abstract, such as a common topic of conversation. These joint interactions within the scanner have been difficult to study given that they require that a person in the scanner interact with another person in a manner that communicates something about an object or goal. As described above, the challenge behind creating a naturalistic social interaction while lying motionless and alone in a noisy tube is not trivial. Through several lines of research, this challenge is beginning to be overcome.

Triadic attention with an alleged partner

One solution to the problem of collecting neuroimaging data during a triadic interaction is to have the participants think they are interacting with another person in a common task. For example, participants are engaged in a trust or decision-making game with an alleged human partner (who they can not see). These studies compare patterns of activation during engagement in a game in which subjects are told they are playing a person (who is not visible) or a computer. When subjects think they are engaged in a collaborative game with a person, greater activation is most consistently found in medial prefrontal cortex (Fukui, et al., 2006; Gallagher, Jack, Roepstorff, & Frith, 2002; Kircher, et al., 2009; Rilling, et al., 2002; Rilling, Sanfey, Aronson, Nystrom, & Cohen, 2004). One interpretation is that, similar to a social interaction, participants need to monitor the other players thoughts and beliefs in relation to their own to achieve a shared goal. This representation of another is only required when playing a human.

Triadic attention with a visible person

Recent studies have come up with novel methods to allow participants to interact face-to-face with another person (either real, virtual, or a video of a real person). Participants are engaged in a joint attention game in which a (visible) experimenter and subject share attention on an object, simulating (or actually creating in some cases) a joint attention experience. These studies (Bristow, Rees, & Frith, 2007; Materna, et al., 2008; Redcay, et al., 2010; Schilbach, et al., 2009; Williams, Waiter, Perra, Perrett, & Whiten, 2005) have identified recruitment of

regions within the medial prefrontal cortex (Bristow, et al., 2007; Redcay, et al., 2010; Schilbach, et al., 2009; Williams, et al., 2005) and posterior superior temporal sulcus (Materna, et al., 2008; Redcay, et al., 2010) while participants share attention on an object with another person. In each study, the design used to elicit joint attention varied widely; however the regions engaged during these joint attention tasks are consistent.

One dimension by which the studies differ is how the participant and 'experimenter' achieve shared attention. In one study (Williams, et al., 2005) joint attention was achieved coincidentally, rather than intentionally. Study participants were instructed to follow a ball that was moving in the lower half of the screen. In the upper half of the screen, a video of a man's face was presented. The man either also followed the ball – leading to an experience of joint attention between the man, subject, and ball – or did not follow the ball (nonjoint attention). In this study joint as compared to nonjoint attention differentially recruited activity in both MPFC as well as posterior cingulate cortex. Thus these regions appear to be engaged during the experience of shared attention even if neither party initiated or responded to a bid for joint attention.

In the remainder of the studies joint attention is achieved by following the experimenter's gaze cue in order to share attention with the experimenter and achieve a goal (Materna, et al., 2008; Redcay, et al., 2010; Schilbach, et al., 2009). Theses studies varied by whether a real or virtual character was used and by the condition used to control for properties of the interaction not relevant to joint attention. In one of these studies (Schilbach, et al., 2009), participants were cued by

a virtual character towards one of 3 squares. When the participant and the virtual character both looked at the square, the square lit up. In the comparison condition the subjects were told to look at a square that was not cued by the virtual character. In this case, the square that the subjects looked at, not the virtual character, lit up. This study identified medial prefrontal cortex as differentially recruited during the joint attention condition as compared to the control (Schilbach, et al., 2009).

A second study presented subjects with a virtual character that would both shift her gaze towards one of five targets on the screen and her iris would change color to match one of the five targets on each trial (Materna, et al., 2008). On joint attention trials, participants were told to look at the same target at which the face was looking. On nonjoint attention trials, participants were told to look at the target that matched the color of the woman's iris. Thus, in both conditions the participant observed a gaze shift; however, only in joint attention trials did the participant use that gaze shift to direct their own attention and share attention with the face image. Activation was seen in right posterior superior temporal cortex in joint as compared to non-joint attention conditions.

Finally, in the third study (Redcay, et al., 2010) participants were engaged in an interactive game with an experimenter in a face-to-face interaction (via a live video feed). The participant was told their goal was to 'catch the mouse'. In the joint attention condition, the experimenter would receive a cue as to which of the four corners of the screen the mouse was 'hiding'. The experimenter would then look at that corner. The subject would detect the experimenter's gaze shift and follow her gaze to the appropriate corner. When the subject also looked to the corner the

mouse was 'caught' and appeared on the screen. In the control (non-joint attention) condition, the subject received the cue as to where the mouse was 'hiding' and simply looked to that corner of the screen in order to 'catch' the mouse. The experimenter opened and closed her eyes at the start of each trial in order to control for the presence of biological motion and make explicit to the subject that the experimenter was not involved in the game. Joint as compared to non-joint attention recruited regions within bilateral posterior superior temporal cortex with the strongest focus of activation in the right posterior superior temporal sulcus as well as a region within dorsal medial prefrontal cortex.

Thus, the posterior STS appears to be recruited not only to use a gaze cue to shift one's attention but also to share attention with another on an object (Materna, et al., 2008; Redcay, et al., 2010). The medial prefrontal cortex also appears to be recruited when attention is shared intentionally between two people and an object (Redcay, et al., 2010; Schilbach, et al., 2009).

Initiating Joint Attention

Until now, we've focused on just one side of a joint attention interaction: that of the responder. However, unlike detecting a gaze shift, monitoring the relationship between one's own attention, another's, and an object is required by both the responder and initiator in a joint attention episode. Thus, a powerful test of the regions important to this fourth component is to examine whether the same regions are engaged during joint attention in both the initiator and responder.

The ability to study the initiation of a joint attention episode poses a significant challenge during fMRI data acquisition for two reasons: (1) as with responding to joint attention, initiating joint attention requires that the subject believe she is in a joint attention episode with another person and (2) the response of the experimenter must be contingent on the subjects, which is unique to joint attention. The above section addressed how studies have overcome challenge 1. Below we will describe how challenge 2 is just beginning to be addressed.

One way to accomplish challenge 2 is to use eye-tracking technology to detect exactly where a participant is looking at all times. In this way, the stimuli can be programmed in such a way that the subject's gaze behavior determines which stimuli are presented to the subject. This method was recently developed (Wilms, et al., 2010) and used in a study of joint attention (Schilbach, et al., 2009). Participants viewed a virtual character but they were told that this character mimicked the behavior of an actual person. The authors used a design in which gaze was either followed (joint attention) or not followed (nonjoint attention) and in which either the subject (self) or experimenter (other) initiated the gaze shift. In the self (initiating) joint attention condition, when the eye-tracking software detected the subject's gaze over one of the four squares, the virtual character's gaze would shift to that block, allowing for a contingent joint attention interaction. Both types of joint attention combined (initiating and responding) as compared to nonjoint attention conditions recruited multiple brain areas including the MPFC.

A second way to overcome the challenge is to develop a set-up in which a real person can interact with a subject face-to-face (via video feed) and in real-time. Recently, we developed such a set-up (Redcay, et al., 2010). During the experiment participants engaged in simple and highly-scripted interactive games with an experimenter. This experiment did not explicitly test joint attention; however numerous episodes of joint attention occurred within the interaction. In the 'live' condition, the experimenter and subject engaged in a real-time, live interaction whereas in the 'recorded' conditions, subjects responded to a recording of the experimenter. Thus, the key difference between these conditions is that only in the live condition are the experimenter's actions directly contingent on the subject's gaze behavior. Comparison of live versus recorded conditions revealed greater activation in a number of brain regions, including the right posterior STS, which is consistent with it's role in monitoring another person's attention and actions in relation to your own, as is required in a joint attention interaction (Redcay, et al., 2010)

Using the same face-to-face set-up, we examined the neural bases underlying initiating joint attention specifically. In this unpublished study, we used a similar design to that described in the Redcay et al., (2010) joint attention study in which participants were instructed to play an interactive game 'catch the mouse' with a live experimenter. In addition to the responding to joint attention and control conditions described above a third condition, initiating joint attention, was included. In this condition, the subject received a clue as to where the mouse was hiding on his screen. Based on the subject's gaze, the experimenter shifted her gaze to that

location and the mouse was then caught. Comparison of both responding to and initiating joint attention conditions with the control condition revealed activation in dMPFC and bilateral pSTS.

MPFC and pSTS are key players in joint attention

In sum, the MPFC and pSTS clearly play a key role in joint attention; however the specific role remains undetermined. Each plays a distinct role in components of joint attention. The MPFC is recruited when someone is perceived as a social partner. The pSTS is recruited to detect a shift in another person's attention. However, both are recruited when participants are required to share attention with another person on an object. These regions are recruited whether or not the participant responds to a gaze cue or initiates a gaze cue. Designs have varied widely and reveal differential involvement of these two regions (described above) but the reason for these differences is not yet clear.

A major question for future research is whether the MPFC and pSTS play distinct roles in the key component of joint attention: namely the intentional, active coordination of attention between two people and an object. One possibility is that both together serve a function during joint attention that can not be isolated to one or another region. Another possibility is that each are engaged during a joint attention context but represent slightly different aspects of that interaction. For example, the MPFC may represent the 'what' of another' attention and the pSTS may represent the 'where'. In other words, the MPFC is recruited to monitor what the other person is attending to in relation to your own attention (e.g. an abstract game

or the properties of an object) while the pSTS is recruited to represent where one's attention is in relation to one's own.

While the specific roles of each region remain unclear, the involvement of the MPFC and pSTS in joint attention in adults is clear. Comparatively less is known about the neural bases of joint attention in infancy.

Neural mechanisms of joint attention in development

As defined above, joint attention depends on (1) detection of a potential social partner (2) detection of a shift in another's attention (3) shifting attention to the appropriate location and (4) monitoring the relation between another person's attention in relation to one's own and a third entity. Similar to the adult literature, research has focused mostly on the first, second, and third components as these are easier to isolate and study in infants. The challenge in infancy is how to measure whether infants are simply orienting to others attention or whether they understand the intention behind another's shift in attention. As Call and Tomasello note (2005), joint attention is not simply looking at the same thing as another person, but rather it is *knowing* both of you are looking at the same thing at the same time (Call & Tomasello, 2005). This active, intentional coordination is a critical component of joint attention behavior. When this *knowing* emerges in development is a matter of debate. Some suggest joint attention does not emerge until infants have the capacity to understand another's intention, which occurs at the end of the first year of life (Tomasello, Carpenter, Call, Behne, & Moll, 2005). Others suggest that joint attention emerges as early as 4 months of age (Grossmann & Johnson,

2010). In the following section we will review both behavioral and neuroscience evidence examining when joint attention emerges and what neural and cognitive systems underlie its emergence.

What can neuroimaging measures provide?

Neuroimaging measures can provide insight into two questions for which behavioral evidence alone may not be sufficient. First, neuroimaging can provide greater sensitivity to ask questions of very young infants who cannot or will not give reliable overt behavioral responses. These measures can reveal discrimination between different stimuli without requiring any measure of overt behavior. Second, neuroimaging measures can provide information about what regions or processes are differentially engaged. In our review, we will pay particular attention as to whether neuroimaging has yet lived up to its potential. Has neuroimaging provided us with anything additional beyond what we know from behavior?

What do we know from behavioral studies?

Full joint attention is thought to emerge between 9-12 months of age (Tomasello, et al., 2005). However, components, or precursors, of this process are present early and develop throughout the first few years of life. For example, infants appear biased to social stimuli from the first minutes of life (Johnson, Dziurawiec, Ellis, & Morton, 1991) and can discriminate faces with direct gaze as compared to averted gaze (Farroni, Csibra, Simion, & Johnson, 2002). As early as 2-3 months of age infants can detect the contingency between their own and another's actions within a social interaction (Murray & Trevarthen, 1985). Thus, there is evidence that

the first component of joint attention, detecting another as a social partner, is present as early as the first few months of life.

By 3-4 months of age, infants are able to use directional cues to shift their attention. Also around this age, infants begin to be able to volitionally disengage attention from their current focus (Colombo, 2001; Hood, Willen, & Driver, 1998; Hunnius, 2007; Richards, 2003). Initially, infants can only shift attention if the object is already within their field of view. The ability to detect less salient directional cues (gaze shift vs. head turn, etc) improves well into the second year of life (Butterworth & Jarrett, 1991, but see Moll & Tomasello, 2004). Nonetheless, the first signs of the ability to shift attention based on another's shift of attention is present by 3-4 months of age. Thus, behaviorally, there is compelling evidence that portions of joint attention are present from very early on, as young as 4 months of age.

Evidence for the fourth component is trickier. When do infants actively monitor their own attention, another's, and an object? When do they *know* they are attending to the same thing as another person? While one may be able to use another's gaze as a directional cue, the understanding of the intention behind that shift is necessary to achieve joint attention. Similarly, to monitor the relation between one's own and another's attention on an object requires an understanding that both parties are intending to attend to that object. Much behavioral evidence suggests this 'knowing' is not present until close to the end of the first year of life. For example, infants at 9 months of age will look just as long at a toy that a woman turned towards with her eyes open or with her eyes closed. By 10-11 months of age

however, infants look longer at the toy only if the woman turned towards it with her

eyes open (Brooks & Meltzoff, 2005). In a separate series of studies, infants were

first habituated to a woman who gazed at either a bear or a ball (Woodward, 2003).

On new object trials the woman looked to the same side of the screen but the object

changed. On new side trials the woman looked to the other side of the screen but the

object remained the same. Only infants 12 months of age looked longer on new

object trials but not new side trials suggesting 12 month olds, but not 7-9 month

olds, are able to represent the goal or intention behind another's gaze shift (i.e. to

look at the object). These findings suggest that by 10-11 months, but not earlier,

infants understand the referential intent of another's gaze shift.

What do we know from neuroimaging measures?

Here we draw on neuroimaging as a possible tool to ask whether there is

evidence for an early representation of the referential intention behind a gaze shift.

The majority of neuroimaging in infancy has been conducted through the use of

event-related potentials (ERPs) (Nelson & McCleery, 2008). Event-related potentials

are recorded from a cap placed over the scalp that contains electrodes (sensors)

that detect electrical activity coming from the brain. Brain cells firing generate this

electrical activity. In the ERP method, the activity is time-locked to the presentation

of a stimulus, allowing for a measure of how a stimulus affects the amplitude of the

signal over time. Amplitude can be positive or negative and both reflect increased

firing. These signals are described as components, which are simply terms that

reflect the time at which a relative peak in amplitude is detected and whether the

activity is positive or negative. For example, the N170 component describes electrical activity that occurs at 170 milliseconds after a stimulus and is negative in amplitude. Many of these components are associated with different processes, such as attention or memory. In this way, ERPs can index whether an infant is paying more or less attention to a stimulus without having to rely on behavior. On the other hand, ERPs provide fairly limited information about where the activity is coming from so it makes comparison to the adult fMRI results difficult. In the next section, we will examine studies that have used the ERP method to examine when infants show evidence for triadic attention.

1. When are infants sensitive to joint attention?

Using ERPs two studies have examined 9 month-old infants' neural sensitivity to joint attention contexts (Senju, Johnson, & Csibra, 2006; Striano, Reid, & Hoehl, 2006). In a novel, live interactive set-up, Striano et al. (2006) showed that 9 month olds process objects differently if they are observed in a joint as compared to non joint attention context. In the joint attention condition, the adult first made eye contact with the infant and then looked at the object on the screen. In the nonjoint attention condition the adult just looked at the object on the screen. ERPs were analyzed during presentation of the objects in either joint or nonjoint attention contexts. A fronto-central negativity (known as the Nc component), which is thought to reflect attentional processing (Courchesne, Ganz, & Norcia, 1981) showed larger amplitudes to the objects which were viewed during joint attention situations than non-joint attention situations (Striano, et al., 2006), suggesting greater attentional

resources are devoted to processing objects when they are in a joint attention context.

In a second study (Senju, et al., 2006), 9-month old infants (and adults) were presented with a screen containing an image of a woman's face. During each trial, an object would appear on the left or right of the screen and then the woman would shift gaze either toward (object-congruent) or away (object-incongruent) from the object. Both infants and adults showed greater negativity in components over occipito-temporal sites to object-*incongruent* gaze shifts. These findings suggest some continuity in the response to object-incongruent gaze shifts between 9 months and adulthood. However, unlike adults, infants showed greater negativity (N200 and N400) in anterior electrode sites during object-congruent gaze shifts. One possible interpretation of the anterior negativity in the infants but not adults is that the neural response in infants' is less specialized than that of adults (Johnson, Grossmann, & Cohen Kadosh, 2009).

Infants show discrimination of joint attention contexts as young as 4 months

As noted above, a major potential contribution of neuroimaging in infancy is to ask when behaviors emerge. Studies have examined even younger ages in an attempt to answer just that. Infant ERP measures reveal that by 4 months of age, infants can detect that a person is looking at an object; process an object more if gazed at by another; and use an other's emotional information to modulate object-processing (review, Striano & Reid, 2006). Four-month old infants show a greater positive slow wave (PSW), which is thought to be a measure of encoding (Nelson,

1996), if an adult's gaze is directed at an object versus not at an object (Hoehl, Reid, Mooney, & Striano, 2008). They show a smaller PSW if they view an object that they had previously seen an adult gaze at, than if they view an object that an adult had not gazed at, suggesting the object had already been encoded during the joint attention episode (Reid, Striano, Kaufman, & Johnson, 2004).

Infants as young as 4-months old also show some evidence that they expect an adult to look towards an object. Specifically, if an object appears and an adult looks to the other side of the screen, infants show a larger Nc component (an index of attentional processing) than if the adult looks towards the object (Hoehl, Reid, et al., 2008). They also show a larger Nc if they view an object which was previously viewed by an adult with a fearful as compared to neutral gaze, suggesting the infant is able to detect an adult's negative emotion and map it onto the object the adult is viewing (Hoehl, Wiese, & Striano, 2008).

In sum, ERP studies suggest sensitivity to joint attention contexts as early as 4 months of age. This early neural sensitivity is intriguing given behavioral results suggesting that infants do not map the intention behind a gaze shift until closer to the end of the first year of life. One possibility is many of the neural results are interpretable in the context of low-level properties. For example, direction of attention may serve as a directional cue to infants, similar to a non-social stimulus such as an arrow. With a directional cue towards an object the infant may process that object more, simply because the infant's own attention is now on the object. Similarly, another person's fearful expression may make the directional cue more

salient or induce an arousal response in the infant allowing for more attentional resources devoted to the object.

On the other hand, ERP studies provide some evidence for continuity in the neural response to joint attention between infants and adults. For example, greater occipito-temporal negativity to an object-incongruent gaze shift is observed in infant ERPs, adult ERPs, and adult fMRI (in the pSTS) (Senju, et al., 2006). In adults this activity likely reflects the fact that an object-incongruent gaze shift violates the observer's expectation of the experimenter's action, thus requiring greater processing of the true intention behind that action. Could infants have such a representation of other's intentions at only 4 months of age? Compelling evidence would be if ERPs were recorded during a similar paradigm to that of Woodward (2003). If infants show neural discrimination between new object (goal) and new side trials this would suggest infants have a neural representation of the goal of the reach. To our knowledge no such study has been conducted.

2. What brain regions are engaged during joint attention?

One way to examine whether the infant brain is sensitive to the intention behind another's gaze shift is to see if the same regions that are recruited in adults (who we know are sensitive to intention) are also recruited in infants during a joint attention episode. Review of the adult literature suggests that the MPFC and pSTS are key players in active and intentional sharing of attention with another person on an object in adults. Are the same regions engaged in infants?

The adult studies relied on fMRI because it is a non-invasive method that has good spatial resolution (i.e. can identify which brain regions are involved at the resolution of about 45 mm^3). However, fMRI is very challenging to use with infants because it requires infants remain almost motionless in a big, noisy tube while focusing attention on a mirror above their head. A relatively new technique, Near Infrared Spectroscopy (NIRS) eliminates a number of these challenges (Lloyd-Fox, Blasi, & Elwell, 2010). NIRS, like fMRI, uses changes in blood volume (or flow) as an index of neuronal activity, which allows for some comparison across methodologies. NIRS works by shining a light across the scalp and measuring the amount of light absorbed. The absorption level will vary based on the volume of hemoglobin in that area, an indirect index of neural activity. The resolution is not as good as fMRI but can still give some degree of localization. Instead of lying in a tube, infants wear a lightweight hat and as a result the images are less affected by infant head motion. Thus, in principle NIRS offers an exciting avenue for direct comparison of neuroimaging results in infants and children.

dMPFC recruited during triadic interaction at 5 months

The one study to examine triadic attention in infancy with NIRS (Grossmann & Johnson, 2010) did so in 5-month old infants (before the age at which infants are thought to understand intentions behind other's actions). They find that a region within the left dorsal medial prefrontal cortex (dMPFC) is selectively recruited during a joint attention episode in which an adult engages an infant in a communicative interaction (by making eye contact) and then shifts attention to an

object. This same selectivity is not seen if the adult either does not look at the infant first or looks away from the object. The study only examined regions within frontal cortex so it is not clear if posterior areas were engaged during this joint attention episode. Nonetheless, engagement of dMPFC suggests that even at 5 months, infants recruit a similar neural region as older infants and adults.

dMPFC and pSTS recruited during dyadic interaction at 4 months

A second NIRS study examined regions within both frontal and posterior areas in 4-month olds (Grossmann, et al., 2008). This study only examined the first component of joint attention, namely engaging in a social interaction in 4-month old infants. Using NIRS, they found that like in adults (Kampe, et al., 2003; Pelphrey, Viola, et al., 2004; Redcay, et al., 2010; Schilbach, et al., 2006), the dMPFC and right pSTS were recruited during engagement in a social interaction. Specifically, sensors over the medial prefrontal cortex and posterior superior temporal cortex show greater activation if a virtual character looked toward the infant and made a communicative expression (raised eyebrows and a smile) than if the character looked away from the infant and also made the same communicative expression.

Continuity in brain regions underlying joint attention between infants and adults

In sum, recent evidence suggests that infants as young as 4-5 months of age and adults recruit similar neural regions to both detect another as a potential social partner and to engage in a joint attention episode. Like adults, infants rely on dorsal medial prefrontal cortex and the posterior superior temporal cortex.

One possible conclusion is that the continuity in neural regions supporting joint attention between infancy and adulthood provides evidence that intention understanding is a robust and very early-emerging phenomenon. If true, this would be a case where neuroimaging measures can provide a more sensitive measure of infant cognition. However, this claim should be taken with caution given that we know the MPFC and pSTS are also recruited during the early components of joint attention: the detection of another as a social partner and perception of gaze direction. Thus, recruitment of these regions in infants does not necessarily imply intention understanding.

An alternative possibility is that these regions may serve as an early-emerging social-communicative system that is initially only partially functional and serves to bias infants towards communicative cues (similar to Grossmann & Johnson, 2007). Detection of these communicative cues is critical for the infant to take maximal advantage of learning opportunities within his or her social world (Csibra & Gergely, 2009). Later development of these systems may allow for the more complex representation of the intention behind one's attention.

Current evidence cannot rule out the possibility that other systems may be critical to joint attention and its development. Evidence from the adult suggests that attention systems are involved in some components of joint attention. Similarly (*as discussed in a chapter within this book*) systems implicated in representing one's own and another's actions (known as the mirror neuron system) may also play a role. In fact, compelling behavioral evidence suggest the infants own experience in

the first year of life provides a necessary foundation for understanding another's actions (Meltzoff & Brooks, 2008; Sommerville, Woodward, & Needham, 2005). The infant NIRS studies do not give a measure of whole-brain activity and the ERP studies do not provide sensitive information on localization of function. Thus, it is possible that other regions and systems (for example, mirror neuron and attention systems) are differentially recruited during the emergence and development of joint attention (for review, Mundy & Newell, 2007 and Mundy, Sullivan, & Mastergeorge, 2009). Replicating the same studies over other regions of cortex would give insight into whether the pattern of neural response across the whole brain is the same in infants as in adults.

Conclusion

Exciting advances in neuroimaging technologies have allowed us to identify the regions that are critical to our ability to engage in a social interaction, follow another's attention, and shift our own attention. Identifying regions that are critical to active, intentional coordination of attention has proven to be a more challenging endeavor. Nevertheless recent creative experimental designs and technological advances (e.g. Redcay, et al., 2010; Schilbach, et al., 2009) have opened a new avenue of research. The first experiments reveal that the MPFC and pSTS are recruited during this process. Recent strides have been made in investigating this question in infants through the use of non-invasive imaging methods including ERP and NIRS. By 4 months, infants are sensitive to a joint attention context and the same regions are recruited as in adults.

Clinical Implications

In the introduction, we posited that the third significant contribution of neuroimaging measures is to provide a baseline by which atypical populations can be compared. Advances in understanding the neuroscience of joint attention may be particularly pertinent to individuals with autism spectrum disorder (ASD), a developmental disorder characterized by impairments in social interaction and communication. ASD individuals show robust and early impairments in joint attention, both in initiating and in responding to others' attention shifts (Charman & al, 1997; Mundy, Sigman, Ungerer, & Sherman, 1986). Given the critical role of joint attention in later language acquisition, emotion regulation, and theory of mind, these early impairments may have a cascading effect on linguistic, emotional, and social development in these children (Charman, 2003; Mundy, et al., 2007; Mundy, et al., 2009). Some preliminary neuroscience work in autism has already revealed that the posterior STS is recruited to the same extent when an ASD subject perceives someone shift gaze either toward or away from an object, suggesting a lack of intention-attribution to the gaze shift (Pelphrey, Morris, & McCarthy, 2005). Other work has shown that, unlike controls, individuals with autism do not show greater ERP responses to faces with direct than averted gaze (Senju, Tojo, Yaguchi, & Hasegawa, 2005), suggesting a lack of attention to another as a potential social partner. No research has yet examined brain differences during active monitoring and sharing of attention to achieve a goal in autism. Future work using naturalistic contexts and explicitly testing what brain regions are recruited during active and

intentional triadic attention holds promise to uncover the mechanism underlying

this early and fundamental impairment in ASD.

References

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Dev, 62*(5), 875-890.

Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: inferential processes versus action simulation. *Curr Biol, 17*(24), 2117-2121.

Bristow, D., Rees, G., & Frith, C. D. (2007). Social interaction modifies neural response to gaze shifts. *Soc Cogn Affect Neurosci, 2*(1), 52-61.

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Dev Sci, 8*(6), 535-543.

Butterworth, G., & Jarrett, N. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology, 9*(1), 55-72.

Call, J., & Tomasello, M. (2005). What Chimpanzees Know about Seeing, Revisited: An Explanation of the Third Kind. In N. Eilan, C. Hoerl, T. McCormack & J. Roessler (Eds.), *Joint Attention: Communication and Other Minds*. Oxford: Clarendon Press.

Charman, T. (2003). Why is joint attention a pivotal skill in autism? *Philos Trans R Soc Lond B Biol Sci, 358*(1430), 315-324.

Charman, T., & al, e. (1997). Infants with autism: an investigation of empathy, pretend play, joint attention, and imitation. *Developmental Psychology, 33*, 781-789.

Colombo, J. (2001). The development of visual attention in infancy. *Annu Rev Psychol, 52*, 337-367.

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., et al. (1998). A common network of functional areas for attention and eye movements. *Neuron, 21*(4), 761-773.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci, 3*(3), 201-215.

Courchesne, E., Ganz, L., & Norcia, A. M. (1981). Event-related brain potentials to human faces in infants. *Child Development, 52*(3).

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends Cogn Sci, 13*(4), 148-153.

Engell, A. D., Nummenmaa, L., Oosterhof, N. N., Henson, R. N., Haxby, J. V., & Calder, A. J. (2010). Differential activation of frontoparietal attention networks by social and symbolic spatial cues. *Social Cognitive and Affective Neuroscience, Epub Ahead of Print*.

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proc Natl Acad Sci U S A, 99*(14), 9602-9605.

Fukui, H., Murai, T., Shinozaki, J., Aso, T., Fukuyama, H., Hayashi, T., et al. (2006). The neural basis of social tactics: An fMRI study. *Neuroimage, 32*(2), 913-920.

Gallagher, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *Neuroimage, 16*(3 Pt 1), 814-821.

Grossmann, T., & Johnson, M. H. (2007). The development of the social brain in human infancy. *Eur J Neurosci, 25*(4), 909-919.

Grossmann, T., & Johnson, M. H. (2010). Selective prefrontal cortex responses to joint attention in early infancy. *Biol Lett.*

Grossmann, T., Johnson, M. H., Lloyd-Fox, S., Blasi, A., Deligianni, F., Elwell, C., et al. (2008). Early cortical specialization for face-to-face communication in human infants. *Proc Biol Sci, 275*(1653), 2803-2811.

Harris, L. T., Todorov, A., & Fiske, S. T. (2005). Attributions on the brain: neuro-imaging dispositional inferences, beyond theory of mind. *Neuroimage, 28*(4), 763-769.

Hasselmo, M. E., Rolls, E. T., Baylis, G. C., & Nalwa, V. (1989). Object-centered encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey. *Exp Brain Res, 75*(2), 417-429.

Hoehl, S., Reid, V., Mooney, J., & Striano, T. (2008). What are you looking at? Infants' neural processing of an adult's object-directed eye gaze. *Dev Sci, 11*(1), 10-16.

Hoehl, S., Wiese, L., & Striano, T. (2008). Young infants' neural processing of objects is affected by eye gaze direction and emotional expression. *PLoS One, 3*(6), e2389.

Hood, B. M., Willen, D., & Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychological Science, 9*(2), 131-134.

Hooker, C. I., Paller, K. A., Gitelman, D. R., Parrish, T. B., Mesulam, M. M., & Reber, P. J. (2003). Brain networks for analyzing eye gaze. *Brain Res Cogn Brain Res, 17*(2), 406-418.

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nat Neurosci, 3*(3), 284-291.

Hunnius, S. (2007). The early development of visual attention and its implications for social and cognitive development. *Prog Brain Res, 164*, 187-209.

Jellema, T., Baker, C. I., Wicker, B., & Perrett, D. I. (2000). Neural representation for the perception of the intentionality of actions. *Brain Cogn, 44*(2), 280-302.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition, 40*(1-2), 1-19.

Johnson, M. H., Grossmann, T., & Cohen Kadosh, K. (2009). Mapping functional brain development: Building a social brain through interactive specialization. *Dev Psychol, 45*(1), 151-159.

Kampe, K. K., Frith, C. D., & Frith, U. (2003). "Hey John": signals conveying communicative intention toward the self activate brain regions associated with "mentalizing," regardless of modality. *J Neurosci, 23*(12), 5258-5263.

Kingstone, A., Tipper, C., Ristic, J., & Ngan, E. (2004). The eyes have it!: an fMRI investigation. *Brain Cogn, 55*(2), 269-271.

Kircher, T., Blumel, I., Marjoram, D., Lataster, T., Krabbendam, L., Weber, J., et al. (2009). Online mentalising investigated with functional MRI. *Neurosci Lett, 454*(3), 176-181.

Kuzmanovic, B., Georgescu, A. L., Eickhoff, S. B., Shah, N. J., Bente, G., Fink, G. R., et al. (2009). Duration matters: dissociating neural correlates of detection and evaluation of social gaze. *Neuroimage, 46*(4), 1154-1163.

Lloyd-Fox, S., Blasi, A., & Elwell, C. E. (2010). Illuminating the developing brain: the past, present and future of functional near infrared spectroscopy. *Neurosci Biobehav Rev, 34*(3), 269-284.

Materna, S., Dicke, P. W., & Thier, P. (2008). Dissociable roles of the superior temporal sulcus and the intraparietal sulcus in joint attention: a functional magnetic resonance imaging study. *J Cogn Neurosci, 20*(1), 108-119.

Meltzoff, A. N., & Brooks, R. (2008). Self-experience as a mechanism for learning about others: a training study in social cognition. *Dev Psychol, 44*(5), 1257-1265.

Mitchell, J. P., Heatherton, T. F., & Macrae, C. N. (2002). Distinct neural systems subserve person and object knowledge. *Proc Natl Acad Sci U S A, 99*(23), 15238-15243.

Mitchell, J. P., Neil Macrae, C., & Banaji, M. R. (2005). Forming impressions of people versus inanimate objects: social-cognitive processing in the medial prefrontal cortex. *Neuroimage, 26*(1), 251-257.

Moll, H., & Tomasello, M. (2004). 12- and 18-month-old infants follow gaze to spaces behind barriers. *Dev Sci, 7*(1), F1-9.

Morales, M., Mundy, P., Crowson, M., Neal, A. R., & Delgado, C. (2005). Individual differences in infant joint attention skills, joint attention, and emotion regulation. *International Journal of Behavioral Development, 29*(3), 259-263.

Morales, M., Mundy, P., & Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior & Development, 21*(2), 373-377.

Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., & Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child Dev, 78*(3), 938-954.

Mundy, P., & Newell, L. (2007). Attention, Joint Attention, and Social Cognition. *Curr Dir Psychol Sci, 16*(5), 269-274.

Mundy, P., Sigman, M., Ungerer, J., & Sherman, T. (1986). Defining the social deficits of autism: the contribution of non-verbal communication measures. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 27*(5), 657-669.

Mundy, P., Sullivan, L., & Mastergeorge, A. M. (2009). A parallel and distributed-processing model of joint attention, social cognition and autism. *Autism Res, 2*(1), 2-21.

Murray, L., & Trevarthen, C. (1985). Emotional regulation of interactions between two-month-olds and their mothers. In T. M. Field & N. A. Fox (Eds.), *Social perception in Infancy* (pp. 177-198). Norwood, NJ: ABlex.

Nelson, C. A. (1996). Electrophysiological correlates of memory development in the first year of life. In H. Reese & M. Franzen (Eds.), *Biological and neuropsychological mechanisms: Life span developmental psychology* (pp. 95-131). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Nelson, C. A., 3rd, & McCleery, J. P. (2008). Use of event-related potentials in the study of typical and atypical development. *J Am Acad Child Adolesc Psychiatry, 47*(11), 1252-1261.

Nelson, P. B., Adamson, L. B., & Bakeman, R. (2008). Toddlers' joint engagement experience facilitates preschoolers' acquisition of theory of mind. *Dev Sci, 11*(6), 847-852.

Nobre, A. C., Gitelman, D. R., Dias, E. C., & Mesulam, M. M. (2000). Covert visual spatial orienting and saccades: overlapping neural systems. *Neuroimage, 11*(3), 210-216.

Pelphrey, K. A., Morris, J. P., & McCarthy, G. (2004). Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *J Cogn Neurosci, 16*(10), 1706-1716.

Pelphrey, K. A., Morris, J. P., & McCarthy, G. (2005). Neural basis of eye gaze processing deficits in autism. *Brain, 128*(Pt 5), 1038-1048.

Pelphrey, K. A., Singerman, J. D., Allison, T., & McCarthy, G. (2003). Brain activation evoked by perception of gaze shifts: the influence of context. *Neuropsychologia, 41*(2), 156-170.

Pelphrey, K. A., Viola, R. J., & McCarthy, G. (2004). When strangers pass: processing of mutual and averted social gaze in the superior temporal sulcus. *Psychol Sci, 15*(9), 598-603.

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philos Trans R Soc Lond B Biol Sci, 335*(1273), 23-30.

Posner, M. I. (1980). Orienting of attention. *Q J Exp Psychol, 32*(1), 3-25.

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *J Neurosci, 18*(6), 2188-2199.

Redcay, E., Dodell-Feder, D., Pearrow, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D., et al. (2010). Live face-to-face interaction during fMRI: a new tool for social cognitive neuroscience. *Neuroimage, 50*(4), 1639-1647.

Reid, V. M., Striano, T., Kaufman, J., & Johnson, M. H. (2004). Eye gaze cueing facilitates neural processing of objects in 4-month-old infants. *Neuroreport, 15*(16), 2553-2555.

Richards, J. E. (2003). The development of visual attention and the brain. In M. de Haan & M. H. Johnson (Eds.), *The cognitive neuroscience of development* (pp. 73-93). New York, NY: Psychology Press.

Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., & Kilts, C. (2002). A neural basis for social cooperation. *Neuron, 35*(2), 395-405.

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). The neural correlates of theory of mind within interpersonal interactions. *Neuroimage, 22*(4), 1694-1703.

Rosen, A. C., Rao, S. M., Caffarra, P., Scaglioni, A., Bobholz, J. A., Woodley, S. J., et al. (1999). Neural basis of endogenous and exogenous spatial orienting. A functional MRI study. *J Cogn Neurosci, 11*(2), 135-152.

Saxe, R., Xiao, D. K., Kovacs, G., Perrett, D. I., & Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia, 42*(11), 1435-1446.

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., et al. (2009). Minds Made for Sharing: Initiating Joint Attention Recruits Reward-related Neurocircuitry. *J Cogn Neurosci*.

Schilbach, L., Wohlschlaeger, A. M., Kraemer, N. C., Newen, A., Shah, N. J., Fink, G. R., et al. (2006). Being with virtual others: Neural correlates of social interaction. *Neuropsychologia, 44*(5), 718-730.

Senju, A., Johnson, M. H., & Csibra, G. (2006). The development and neural basis of referential gaze perception. *Soc Neurosci, 1*(3-4), 220-234.

Senju, A., Tojo, Y., Yaguchi, K., & Hasegawa, T. (2005). Deviant gaze processing in children with autism: an ERP study. *Neuropsychologia, 43*(9), 1297-1306.

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition, 96*(1), B1-11.

Stevens, S. A., West, G. L., Al-Aidroos, N., Weger, U. W., & Pratt, J. (2008). Testing whether gaze cues and arrow cues produce reflexive or volitional shifts of attention. *Psychon Bull Rev, 15*(6), 1148-1153.

Striano, T., & Reid, V. M. (2006). Social cognition in the first year. *Trends Cogn Sci, 10*(10), 471-476.

Striano, T., Reid, V. M., & Hoehl, S. (2006). Neural mechanisms of joint attention in infancy. *Eur J Neurosci, 23*(10), 2819-2823.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav Brain Sci, 28*(5), 675-691; discussion 691-735.

Vaughan Van Hecke, A., Mundy, P. C., Acra, C. F., Block, J. J., Delgado, C. E., Parlade, M. V., et al. (2007). Infant joint attention, temperament, and social competence in preschool children. *Child Dev, 78*(1), 53-69.

Williams, J. H., Waiter, G. D., Perra, O., Perrett, D. I., & Whiten, A. (2005). An fMRI study of joint attention experience. *Neuroimage, 25*(1), 133-140.

Wilms, M., Schilbach, L., Pfeiffer, U., Bente, G., Fink, G. R., & Vogeley, K. (2010). It's in your eyes--using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. *Soc Cogn Affect Neurosci, 5*(1), 98-107.

Woodward, A. (2003). Infants' developing understanding of the link between looker and object. *Developmental Science, 6*(3), 297-311.