NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 25 (2005) 133–140

An fMRI study of joint attention experience

Justin H.G. Williams,[a,b,*] Gordon D. Waiter,[c] Oliver Perra,[a]
David I. Perrett,[b] and Andrew Whiten[b]

[a]*Department of Child Health, University of Aberdeen, Medical School, Foresterhill, Aberdeen, Scotland, UK*
[b]*Centre for Social Learning and Cognitive Evolution, School of Psychology, University of St Andrews, St Andrews, Scotland, UK*
[c]*Department of Radiology, University of Aberdeen, Lilian Sutton Building, Foresterhill, Aberdeen, Scotland, UK*

Received 1 July 2004; revised 13 October 2004; accepted 28 October 2004
Available online 21 January 2005

Although much is now known about eye movement detection, little is known about the higher cognitive processes involved in joint attention. We developed video stimuli which when watched, engender an experience of joint attention in the observer. This allowed us to compare an experience of joint attention to nonjoint attention within an fMRI scanning environment. Joint attention was associated with activity in the ventromedial frontal cortex, the left superior frontal gyrus (BA10), cingulate cortex, and caudate nuclei. The ventromedial frontal cortex has been consistently shown to be activated during mental state attribution tasks. BA10 may serve a cognitive integration function, which in this case seems to utilize a perception–action matching process. The activation we identified in BA10 overlaps with a location of increased grey matter density that we recently found to be associated with autistic spectrum disorder. This study therefore constitutes evidence that the neural substrate of joint attention also serves a mentalizing function. The developmental failure of this substrate in the left anterior frontal lobe may be important in the etiology of autistic spectrum disorder.

Keywords: Joint attention; Joint attention mechanism; fMRI; Social cognition; Autistic spectrum disorder

Introduction

Joint attention is a process whereby two individuals attend to the same object because one is checking, monitoring, or following the focus of attention of the other. It is an early cognitive skill to emerge; infants will look in the direction of other people's head and eye movements at about 6 months of age (Scaife and Bruner, 1975), and by the age of 18 months they will reliably follow the gaze of others (Butterworth and Jarrett, 1991). It may therefore be an important developmental precursor to other social and cognitive abilities.

In particular, joint attention processes may constitute cognitive precursors to higher level mental state attribution such as that concerned with understanding false belief (Baron-Cohen, 1994; Gopnik et al., 1994; Mundy, 1993). In a longitudinal study, joint attention at 20 months positively correlated with measures of theory of mind collected between 3 and 4 years of age (Charman et al., 2001).

A number of lines of evidence also suggest that joint attention may be an important precursor to the development of language abilities: onset consistently precedes the emergence of referential language in the second year of life; the ability to engage attention jointly during naturalistic infant–mother interactions predicts infants' word comprehension and word production (Carpenter et al., 1998); joint attention bids measured between 6 and 18 months make a unique contribution to language development at 30 months (Morales et al., 2000); and gaze following, which is an important component of joint attention, predicts vocabulary between 1 and 2 years (Morales et al., 1998).

Joint attention skills are notably poor in autism and severity of joint attention impairment predicts the outcome of this disorder (Dawson et al., 2002). In one study (Charman et al., 1997), 12 children with symptoms of autism at 20 months (whose diagnostic status was subsequently confirmed) switched gaze less often between a toy and an experimenter to share aspects of a situation with the experimenter. Children with autism are impaired in their comprehension and production of protodeclarative pointing (pointing aimed at sharing interest with another rather than aimed at requesting objects), and they also seem unable to use the direction of gaze of a person to infer what the person wants (Baron-Cohen, 1989, 1995). They are significantly less likely to check the focus of attention of an experimenter in ambiguous situations compared to control groups (Phillips et al., 1992). However, it appears that these children do possess the ability to compute the direction of another person's gaze and thus to understand what a person is looking at

* Corresponding author. Department of Child Health, Royal Aberdeen Children's Hospital, University of Aberdeen School of Medicine, Westburn Road, Aberdeen AB25 2ZG, Scotland, UK. Fax: +44 1224 663658.
E-mail address: Justin.Williams@abdn.ac.uk (J.H.G. Williams).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2004.10.047

(Leekam et al., 1997, 2000). Recently, children with autism have been shown to be able to use eye direction cues to equal advantage to controls (Kylliainen and Hietanen, 2004; Senju et al., 2004) in reducing their reaction times in responding to a lateralized cue. Rather, it seems they fail to spontaneously follow the direction of gaze of another person with eyes and head turned toward an object (Leekam et al., 2000)—they fail to join in with the other person's attentional focus. This experiment has been largely incorporated into the Autism Diagnostic Observation Schedule (Lord et al., 2000), where joint attention items discriminate well between children with autism and those with other clinical problems.

If gaze detection is functioning normally in autism, it may be that it is another component of the joint attention process that is functioning abnormally in autism. As well as being involved with the detection of another individual's attentional direction, joint attention is also concerned with the tendency to switch gaze between an object and a person, and to direct the attention of another (Leekam et al., 1997; Phillips et al., 1992; Swettenham et al., 1998). Notably, the tendency to direct attention and the ability to (spontaneously) monitor attentional direction are both impaired in autism, suggesting that they could both be affected by a common mechanism, possibly that are concerned with integrating them in the formation of a representation of joint attention. Baron-Cohen (1994, 1995) proposed the existence of a modular 'Shared Attention Mechanism' (SAM), which could serve such an integration function, suggesting that this would be an integral part of a 'theory of mind' mechanism.

In examining the neural correlates of joint attention, there is now strong evidence for specialized neural mechanisms serving gaze and eye movement perception (Baron-Cohen et al., 2001; Hoffman and Haxby, 2000; Jellema et al., 2000; Perrett et al., 1989; Puce and Perrett, 2003; Puce et al., 1998; Wicker et al., 1998). At a higher cognitive level of processing, Wicker et al. (2003) investigated the neural correlates of attributing emotion secondary to gaze processing. Yet, to our knowledge no functional neuroimaging paradigm has been used to investigate integrating information from gaze perception with executed gaze direction at the level of forming a representation of joint attention itself. If joint attention is a cognitive precursor to 'theory of mind' capacity, this may involve similar neural substrate such as ventromedial frontal cortex. In this study, we developed video clips (Fig. 1) that engendered an experience of joint attention in observers who watched them. The joint attention condition was contrasted with a very similar but nonjoint attention condition, where the model's gaze moved equally as often but was always directed elsewhere. This has allowed us to investigate the neural substrate of joint attention at the level concerned with integrating gaze-direction information from the self and another individual.

Methods

Materials

We hypothesized that there would be neural structures more active during joint attention than during a control situation when attention was directed to differing locations. To create these situations, we constructed a number of video clips by positioning a model behind a silk screen, with the video frame showing their head in the top two thirds and the screen in the bottom third. A red dot projected onto the screen moved horizontally and randomly between four points spaced evenly apart. The model watched the dot, moving their head as well as their eyes when they shifted their gaze. As the model was following the dot, there was always a slight delay between the dot movement and the model's gaze, such that the model's gaze was not a useful cue in predicting the movement of the spot. We constructed eight video clips in this manner where the exact timing and sequences of dot moves were different. The video clips were then processed using image software and a plain background replaced the silk screen. Red dots were then pasted back onto the picture in one of two ways. In the first condition, they were pasted so as to be in the line of the model's gaze and the video showed the model to be tracking the object with their gaze as it moved about the screen. In the second condition, they were pasted to be in different positions so that the video clip showed the gaze of the model to be discordant with that of the dot position, and they were seen to be looking elsewhere as the dot moved. The time delay between dot movement and head movement remained equivalent between conditions. There were four clips for each of the two conditions. Each clip lasted 30 s and contained 10 attentional shifts with different orderings of position change and slightly different timing.

Experimental design and task

The clips were shown to participants lying in the scanner. The angle subtended as a result of the distance between the model's eyes and the red dot was about 7°, so that as subjects watched the dot, the gaze direction of the model was easily perceived. They were asked simply to watch the moving dot and not told anything

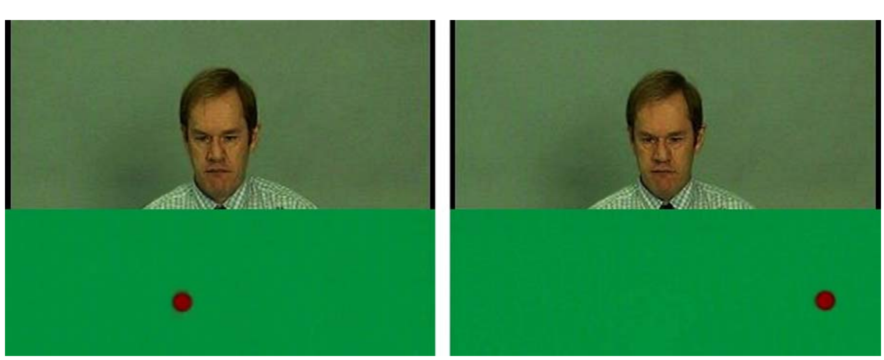

Fig. 1. Still images taken from video clips used as stimuli to create joint attention (left) and nonjoint attention (right). The viewer experiences joint attention while watching the spot in the picture on the left, but not while watching that on the right.