NeuroReport 6, 1741–1746 (1995)

Modeling other minds

Vinod Goel, Jordan Grafman,[CA] Norihiro Sadato[1] and Mark Hallett[1]

Cognitive Neuroscience and [1]Human Motor Control Sections, NIH/NINDS/MNB, Building 10, Room 5S209, 10 Center Drive MSC 1440, Bethesda, MD 20892-1440, USA

[CA]Corresponding Author

NINE normal volunteers performed a 'theory of mind' task while their regional brain blood flow pattern was recorded using the PET [^{15}O]H_2O technique. Control conditions induced subjects to attend to the visual and semantic attributes of known objects. In a third condition, subjects had to infer the function of an unfamiliar object from its form. In the 'theory of mind' condition, subjects had to infer function based on the form of both familiar and unfamiliar objects and in addition, model the knowledge and rationality of another mind about the function of these objects. Performance during the 'theory of mind' condition evoked the activation of a distributed set of neural networks with prominent activation of the left medial frontal lobe (Brodmann area 9) and left temporal lobe (Brodmann areas 21, 39/19, 38). This result suggests that when inferential reasoning depends on constructing a mental model about the beliefs and intentions of others, the participation of the prefrontal cortex is required. When access to such knowledge is affected by central nervous system dysfunction, such as that found in autism, modeling other minds may prove difficult.

Key words: PET; Theory of mind; Object recognition; Autism

Introduction

There is evidence from the cognitive, developmental and philosophical literature to suggest that the human ability to predict with reasonable accuracy and speed the thoughts and actions of others appears early in development and is critical for our mental, physical and social well-being.[1,2] Two types of accounts have been advanced to explain this ability. One account we have an implicit theory about the meaning and logic of propositional attitudes (i.e. beliefs, desires, hopes, fears, etc.) that allows us to attribute them to others and draw appropriate inferences.[1] On another account we engage in a simulation process whereby we make decisions involving the behavior of others in a pretended or simulated context.[3] We construct a mental model of the knowledge states, characteristics, and goals of the other agent, draw the inference ourselves, and then attribute the conclusion to the other agent (i.e. we assume that the other agent will arrive at the same conclusion as we do).

There is some neurophysiological evidence that speaks to the localization of 'theory of mind' reasoning functions. This ability can be selectively impaired by biological disorders such as autism in which both frontal and temporal lobe pathology has been reported.[4–6] Given that autistic children display many of the classic signs of frontal lobe dysfunction, some authors have suggested that our ability to model other minds may be localized in the frontal lobes.[7,8] This suggestion is consistent with reports of impaired social judgments in adult patients with lesions in the ventromedial prefrontal cortex.[9,10]

Recently Baron-Cohen *et al*[7] presented results from a SPECT study that required normal volunteers to identify mental state terms (e.g. think, know, deceive, want, etc.). The control conditions were body-related (e.g. face, eyes, stomach, etc.) and non-mind and non-body (e.g. business, nation, cover, etc.) terms. They reported that the orbito-frontal cortex was significantly more active during the recognition of mind-terms than adjacent frontal areas and the effect was strongest for the right hemisphere.

In this paper we present evidence from a PET ([^{15}O]H_2O) regional cerebral blood flow study that supports the general claim of frontal lobe involvement in modeling other minds. Our experimental design had four main conditions which exposed subjects to familiar and unfamiliar objects, but required them to make distinctive decisions in each condition based on (i) visual perception/shape description, (ii) memory retrieval, (iii) inference from form to function and (iv) inference from form to function involving modeling the knowledge and rationality of another person's mind.

Materials and Methods

Subjects: Ten students (five male, five female), with a mean age of 24.7 (S = 6.18) were recruited from a local university. All subjects were right handed and had normal vision and a similar level of education. One female subject was dropped due to a scan abnormality.

Stimuli: The stimuli consisted of Grey-scale pictures (normal or canonical views) of various kinds of man-made artifacts. Three hundred pictures of artifacts, from 10 different categories (food preparation, food serving, agriculture, hunting and fishing, personal care and adornments, household effects, transportation, manufacturing tools, toys and games, and muscial instruments), were collected, scanned and scaled to a uniform size (50 000 pixels). One hundred and fifty of the artifacts were familiar, easily recognized objects from our time and culture (e.g. hair dryer, blender, automobile, etc.). They were collected from various shopping catalogues. The other 150 artifacts were from the same categories but from pre-fifteenth century Eskimo and North American Indian cultures. The majority were collected from Miles.[11] Subjects should be familiar with the first set of artifacts, but not the second set.

Experiment design: The stimuli were presented in four different conditions (which will be referred to as C1–C4) each repeated once (C5–C8), for a total of eight conditions. In the baseline condition (C1 and C5), 150 of the stimuli (75 familiar, 75 unfamiliar) were presented in each condition and subjects were instructed to respond 'yes' if the depicted artifact is elongated along the principle axis, otherwise respond 'no'. This requires the following cognitive processes: visual perception and analysis, shape discrimination, simple decision, and motor response.

For memory retrieval (C2 and C6), the modern, familiar stimuli were presented (75 in each condition) and subjects were instructed, in C2, to respond 'yes' if the artifact is used for food preparation, otherwise respond 'no' (taking care to differentiate food preparation from food procurement and food serving). In C6 they were required to respond 'yes' if the artifact is used for personal care and adornment, otherwise respond 'no'. This condition requires all of the steps in the baseline condition plus successful retrieval of functional information about the artifacts from long-term memory.

To test simple inference (C3 and C7), the unfamiliar stimuli were presented (75 in each condition) and subjects were instructed in C3 to respond 'yes' if the artifact is used for food preparation, otherwise respond 'no' (taking care to differentiate food preparation from food procurement and food serving). In C7 they were required to respond 'yes' if the artifact is used for personal care and adornment, otherwise respond 'no'. This condition requires all of the steps in the baseline conditions. However, when subjects try to retrieve functional level information about the artifacts from memory, the search will fail in most cases. Subjects will then be required to infer the functional information based on such pictorial cues as shape, texture, material, and perhaps size.

In the theory of mind conditions (C4 and C8), the full set of 150 stimuli was presented in each of these conditions and subjects were instructed to respond 'yes' if they thought that someone with a background knowledge of Christopher Columbus could infer the function of the artifact (that is correctly classify it as belonging to one of the 10 categories), otherwise respond 'no'. Subjects were told to assume that Columbus would have access to the same picture and would not be allowed to touch the object or watch it in operation. This condition requires all of the above three steps — baseline, retrieval of functional information from long-term memory (for the familiar objects), an inference of function based on shape, texture, material, and perhaps size (for the unfamiliar objects) — plus an inference involving the knowledge states and mind set of a 15th century European (Christopher Columbus) and a judgment of whether the individual would be familiar with the function of a particular artifact, and whether they could infer that function from the pictorial information.

The presentation of the stimuli and scan began only after subjects indicated that they had read and understood the instructions. Subjects responded 'yes' or 'no' by pressing response buttons with their thumbs.

The stimuli for each condition were randomized. The stimuli appeared on the center of the screen and remained there until the subject responded. If no response was forthcoming within 5 s, the next stimulus appeared. The interstimulus interval for conditions C1, C5, C4, C8 was 0 ms, and 0.5 ms for conditions C2, C3, C6, and C7. (The interstimulus interval was introduced in these latter conditions to ensure that the 75 stimulus presentations would last beyond the duration of the scan. It was not needed in the former conditions which contained 150 stimulus presentations.) The order of presentation of the eight conditions was counterbalanced across subjects.

PET procedure: PET scanning was performed with a Scanditronix PC2048-15B (Uppsala, Sweden) which collected 15 contiguous planes with an in-plane resolution of 6.5 mm full-width half-maximum after reconstruction, and with a center-to-center distance of 6.5 mm, covering 97.5 mm in axial direction. Matrix size and pixel size of the reconstructed images were 128 × 128 and 2 mm, respectively. A transmission scan was obtained with a rotating Germanium-68 source. Based on the reconstructed transmission images, the position of the head was set to cover the entire frontal lobes.

Reconstructed images were obtained by summing the activity during the 60 s period following the first detection of an increase in cerebral radioactivity after the i.v. bolus injection of 37 mCi of $[^{15}O]H_2O$. No arterial blood sampling was performed and thus the images collected are those of tissue activity. Tissue activity recorded by this method has been shown to be linearly related to rCBF.[12,13]

Each subject underwent eight consecutive scans at 12 min intervals. The stimulus presentation began 30 s prior to the injection and continued until all stimuli were exhausted. This period varied from 100–180 s, depending on the condition and the subject. The PET scan sampled performance in each condition only for a 60 s period beginning approximately 30 s after task onset.

Data analysis was performed using SPM software (MRC Cyclotron Unit, UK) in PROMATLAB (Mathworks, Natick, MA) using ANALYZE image display software (BRU, Mayo Foundation, Rochester, MN). The data from each subject were first standardized for brain size and shape and reconstructed parallel to the intercommissural line.[14–16] Each image was smoothed to account for the variation in normal gyral anatomy using a Gaussian filter (FWHMx × FWHMy × FWHMz = 20 × 20 × 12 mm). In the stereotaxic standard space, each voxel was 2 × 2 × 4 mm in size. The effect of global differences in rCBF between scans removed using an analysis of covariance.[16]

Planned linear comparisons of the adjusted mean images followed. All image analyses were performed on a pixel by pixel basis. Comparisons of the difference in condition means were made by the t-statistic using the adjusted pixel error variances for each condition estimated from the analysis of covariance. The value of t for each pixel in each comparison was then transformed to a normal standard distribution (z values) which was independent of the degree of freedom of the error. The resulting set of z values constituted a statistical parametric map (SPM).[17]

To identify the cortical areas activated with different conditions, six linear comparisons were performed: C2–C1, C3–C1, C4–C1, C3–C2, C4–C2, and C4–C3. All regions reported as being significantly activated exceeded the $p < 0.05$ level of significance with Bonferronl-like correction for repeated measurements.[17]

Results

Cognitive results (based on 10 subjects): Subject reaction times (RTs) and response accuracy were measured. Reaction times for the four conditions are noted in Table 1. The main effect of conditions was significant (F (3, 54) = 26.9, p = 0.0001). The general trend is consistent with our working hypothesis that each condition from baseline to theory of mind builds upon the previous one and requires an additional degree of processing. However, the all artifact RTs for the baseline and memory retrieval conditions seem to be reversed. If we analyze the RTs of the familiar and unfamiliar objects separately, we find (consistent with our hypothesis) that the RTs for the unfamiliar objects in the baseline condition were significantly less than the RTs for both the simple and theory of mind conditions (t (9) = 4.6, p = 0.001 and t (9) = 3.4, p = 0.007 respectively). However, the familiar artifacts in the baseline condition had a significantly higher RT associated with them than the unfamiliar objects (t (9) = 3.8, p = 0.004), and the RTs for the familiar objects in the baseline condition were greater than the RTs in the memory retrieval condition (t (9) = 2.9, p = 0.017). It could be that the shapes of the modern, familiar objects were a little more complex than that of the unfamiliar objects and thus judgments were more difficult and took longer. It could also be that, despite instructions, subjects could not help but perform functional processing of the familiar objects in the baseline condition. Overall, the mean RTs for the unfamiliar objects were significantly greater than those of familiar objects (see Table 1), as one would expect (t (9) = 3.3, p = 0.01).

There was no significant main effect of response type (i.e. whether subjects responded 'yes' or 'no') (F (1, 54) = 0.012, p = 0.91), nor any significant interaction between response type and conditions (F (3, 54) = 0.54, p = 0.65). There was, however, a significant main effect of familiarity of artifacts (F (1,18) = 25.5, p = 0.0001) as expected, and a significant interaction between the RTs for yes/no judgments and the RTs for familiar/unfamiliar objects (F (1,18) = 4.38, p = 0.05; see Table 2). Subjects took

Table 1. Mean RTs (ms) by condition

	C1 & C5 baseline	C2 & C6 memory retrieval	C3 & C7 simple inference	C4 & C8 theory of mind	All conditions
Familar artifacts	1688	1315	N/A	3022	2008
Unfamiliar artifacts	1490	N/A	2496	3052	2346
All artifacts	1589	1315	2496	3037	2177

Table 2. Mean RTs (ms) by response and familiarity

	Familiar artifacts	Unfamiliar artifacts
'Yes' response	2132	2424
'No' response	1947	2651

longer to respond 'yes' to familiar objects than to respond 'no', and longer to respond 'no' to unfamiliar objects than to respond 'yes'.

Response accuracy was calculated for the memory retrieval and simple inference conditions. Subjects correctly retrieved the functional information for 79.0% of the familiar objects and and were able to correctly infer the function of 40.8% of the unfamiliar objects.

PET results: Subtraction of the visual perception/shape discrimination baseline condition activation from the memory condition activation revealed that only a small part of the right precuneus remained active (probably concerned with visual processing and imagery) suggesting that the baseline condition also required the activation of structures involved in memory processing (see Table 1). Subtraction of the memory retrieval condition activation from the simple inference condition activation revealed activation only in the cerebellar vermis. Subtraction of the visual perception/shape discrimination baseline condition from the simple inference condition activation revealed activation of the cuneus, precuneus and pons. The cuneus and precuneus activation can be attributed to the intensive visual inspection of the unfamiliar stimulus in an attempt to impute meaning. This was a somewhat surprising finding as we had expected frontal lobe activation in both conditions requiring inferential reasoning.

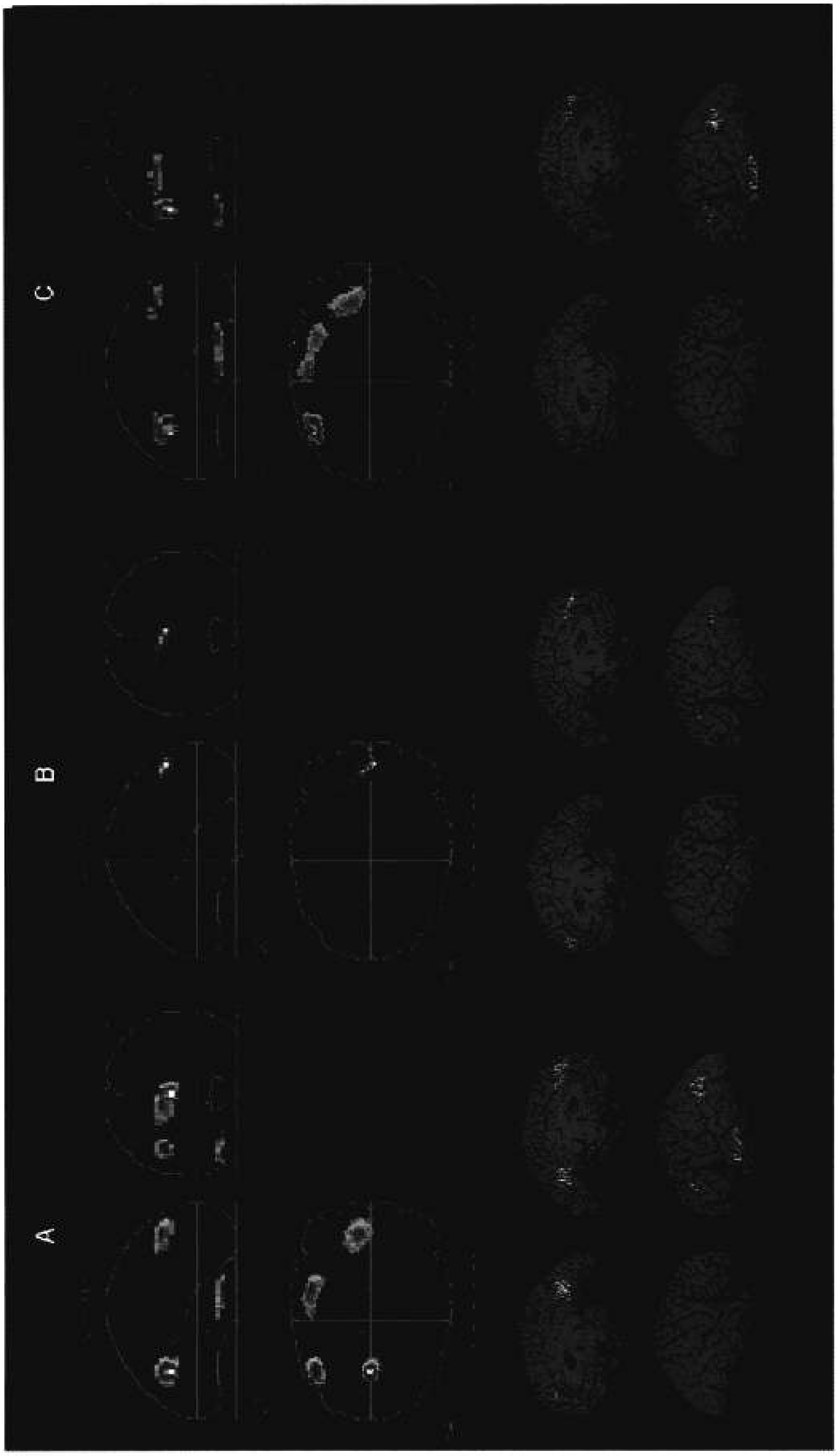
The patterns of brain activation associated with the theory of mind condition (C4–C1, C4–C2, and C4–C3) are depicted in Figure 1 and Table 3. As can be seen in Figure 1, when the activation seen in the three other conditions is independently subtracted from activation obtained in the theory of mind mental inference condition, there always remained a selective activation of the left medial prefrontal cortex and, in one subtraction, the middle lateral prefrontal cortex along with various regions of the left temporal gyrus. In particular, subtracting the activation in the simple inference condition from the activation elicited in the theory of mind condition resulted in a distributed neural network that included the left posterior temporal lobe, left anterior temporal lobe, and left medial frontal lobe. We suggest that the medial and lateral prefrontal cortex subserve the high-level cognitive processes concerned with establishing a mental model whereas the semantic and nominal information required by the model are stored in various sectors of the left temporal cortex.

Discussion

Our results are broadly consistent with the results of lesion studies and other imaging studies. In particular, Baron-Cohen *et al*[7] reported involvement of the right frontal orbito-medial cortex in their task requiring the recognition of mental state terms. We report involvement of the medial frontal lobe (with left predominance), although the selected regions of

Table 3. Location and characteristics of the brain regions that remained significantly active after each hierarchical subtraction of conditions

Location (Brodmann area)	Talairach coordinates			Z-score	%ΔCBF
	X	Y	Z		
Memory retrieval — baseline					
Precuneus (31)	2	−62	20	3.9	2.0
Simple inference — baseline					
Precuneus (31)	2	−62	20	4.4	2.4
Rt cuneus (18)	18	−64	16	4.3	3.5
Pons	−8	−34	−20	4.0	3.8
Simple inference — memory retrieval					
Vermis	4	−84	−24	3.8	7.7
Theory of mind — baseline					
Precuneus (31)	2	−62	20	4.4	3.1
Posterior Lt inferior parietal lobule (39)	−42	−62	24	5.0	3.9
Lt medial frontal gyrus (9)	−6	46	28	4.8	4.3
Lt middle temporal gyrus (21)	−46	2	−20	4.8	5.3
Midbrain	−6	−32	−8	3.7	2.5
Theory of mind — memory retrieval					
Lt medial frontal gyrus (9)	4	52	24	4.3	3.8
Lt middle frontal gyrus (9)	−20	34	32	3.7	3.1
Lt middle temporal gyrus (39/19)	−42	−70	20	3.6	2.5
Theory of mind — simple inference					
Lt middle temporal gyrus (39/19)	−44	−64	20	5.0	3.6
Lt middle temporal gyrus (21)	−48	−16	−16	4.3	4.0
Lt superior temporal gyrus (38)	−44	14	−16	4.3	5.2
Lt medial frontal gyrus (9)	−12	38	32	4.3	4.0


A
B
C

interest of Baron-Cohen *et al* did not encompass the frontal and temporal regions activated in our task: these areas may also have been activated in their task. However, the reason we did not see activation of the right frontal orbito-medial cortex may be due to differences in tasks: there was no inference requirement in the other study. Our task specifically asked subjects to draw inferences based on other people's knowledge states and rationality.

Certain 'planning' or 'executive control' type tasks such as the Tower of London and the Wisconsin Card Sort task result in left mesial and dorsolateral frontal lobe activation in SPECT studies.[18,19] Our results suggest that reasoning involving the modeling of other minds overlaps with this cortical region. Interestingly, some authors have recently begun to explore the relationship between executive functions and modeling of other minds.[8]

One may object that the effect we are attributing to the theory of mind condition is nothing more than an increasing level of processing effect[20] or an increasing complexity effect. The response time data certainly suggest increased processing in the theory of mind condition compared with other conditions. Likewise, the response time data also suggest increased processing between the baseline and simple inference condition, but there is no accompanying frontal and temporal activation. The activation during the theory of mind condition must therefore also have something to do with the type of processing involved, either the simulation requirement or the knowledge of propositional attitudes.

One may also object that subjects were not imagining the mind of Columbus at all, but just classifying objects as new or old, or familiar or unfamiliar. Perhaps they decided that Columbus could not infer the function of modern artifacts but could infer the function of the artifacts from the North America Indian and Eskimo cultures. Again, such an interpretation is not consistent with the cognitive data. If they used such a strategy there is no reason for the reaction times in the theory of mind condition to be significantly higher than in the other conditions. Furthermore, subjects' responses do not break down along the lines of familiar and unfamiliar objects as such a strategy would predict.

In the subtraction of baseline activation from memory retrieval activation, only a small part of the right precuneus emerges as an area of activation. Significant temporal or hippocampal activity was absent. This is consistent with the behavioral results suggesting that the baseline condition may have subsumed the memory retrieval condition.

The simple inference condition did not result in frontal activation. As we know from other imaging studies that have elicited frontal lobe activation,[18,19] such activation is not restricted to inferences involving the modeling of other minds. Our results indicate, however, that the frontal lobes are selectively activated only during certain types of inferential reasoning tasks.

Conclusions

We have presented results from a study requiring subjects to draw inferences of object function based upon information available in gray scale pictures (shape, size, texture, etc.). In one condition subjects were required to model the background knowledge and rationality of another individual as a prerequisite to drawing the inference. The results indicate that inferences requiring the modeling of other minds implicate the frontal lobes, in particular, the left medial cortex (Brodmann area 9). Inferences based on stimulus structure alone can be made without the participation of the frontal lobes.

References

1. Astington JW, Harris PL and Olson DR (eds.) *Developing Theories of Mind.* Cambridge: Cambridge University Press, 1988.
2. Goel V. *Sketches of Thought.* Cambridge, MA: MIT Press, 1995.
3. Gordon RM *Mind Lang* **7**, 11–34 (1992).
4. Leslie AM. The theory of mind impairment in autism: evidence for a modular mechanism of development. In: Whiten A, (Ed.) *Natural Theories of Mind,* Oxford: Blackwell, 1991.
5. Baron-Cohen S. *Psychiatr Clin N Am* **14**, 33–51 (1991).
6. Baron-Cohen S. Tager-Flusberg H and Cohen DJ, ed. *Understanding Other Minds: Perspectives from Autism.* Oxford: Oxford University Press, 1993.
7. Baron-Cohen S, *et al. Br J Psychiatry* **165**, 640–649 (1994).
8. Ozonoff S. Executive functions in autism. In: Schopler E and Mesibov G (eds). *Learning and Cognition in Autism.* New York: Plenum, in press.
9. Eslinger PJ and Damasio AR. *Neurology* **35**, 1731–1741 (1985).
10. Stuss DT and Benson DF. *The Frontal Lobes.* NY: Raven Press, 1986.
11. Miles C. *Indian and Eskimo Artifacts of North America.* Chicago: Henry Regnery Co., 1963.
12. Fox PT *et al. J Cerebr Blood Flow Metab* **4**, 329–333 (1984).
13. Fox PT and Mintun MA. *J Nucl Med* **30**, 141–149 (1989).
14. Talairach J and Tournoux, P. *Co-Planar Stereotaxic Atlas of the Human Brain.* New York: Thieme, 1988.
15. Friston KJ *et al. J Cerebr Blood Flow Metab* **9**, 690–695 (1989).
16. Friston KJ *et al. J Cerebr Blood Flow Metab* **10**, 458–466 (1990).
17. Friston KJ *et al. J Cerebr Blood Flow Metab* **11**, 690–699 (1991).
18. Karim R *et al. Arch Neurol* **50**, 636–642 (1993).
19. Andreasen NC *et al. Arch Gen Psychiatry* **49**, 943–958 (1992).
20. Kapur S. *et al. Proc Natl Acad Sci USA* **91**, 2008–2011 (1994).

Received 24 May 1995; accepted 20 June 1995