In (2000) S. Baron-Cohen, H. Tager-Flusberg & D. Cohen (Eds.), *Understanding other minds: perspectives from developmental cognitive neuroscience - 2nd edition* (pp.306-333). Oxford: Oxford University Press.

Cerebral Lateralization and Theory of Mind

Hiram Brownell[1,2], Richard Griffin[1,3,4]
Ellen Winner[1,3], Ori Friedman[1], & Francesca Happé[5]

(The order of the first two authors is alphabetical and reflects equal contributions.)

[1] Department of Psychology, Boston College, Chestnut Hill, MA
[2] Harold Goodglass Aphasia Research Center, Department of Neurology, Boston University School of Medicine, Boston, MA
[3] Harvard Project Zero, Harvard University, Cambridge, MA
[4] The Center for Cognitive Studies, Tufts University, Medford, MA
[5] Social, Genetic, and Developmental Psychiatry Research Centre, Institute of Psychiatry, London, UK

Acknowledgments

Preparation of this chapter was supported in part by grants 2R01 NS 27894 and DC 00081.We thank Daniel Dennett, Marcel Kinsbourne, Jeff Stewart, Andrew Stringfellow, and Chris Westbury for many suggestions and discussions that have improved this chapter greatly.

Addresses for correspondence:

Richard Griffin
Department of Experimental Psychology
Cambridge University
Cambridge, UK CB2 3EB
office phone: +44 (1223) 746060
email: Rick@post.harvard.edu

Hiram Brownell
Chair, Department of Psychology
Boston College
McGuinn Hall
Chestnut Hill, MA 02126
Email: brownelh@bc.edu

Cerebral Lateralization and Theory of Mind

Introduction

Theory of Mind (ToM), predicting and explaining behavior in terms of mental states, represents a crucial human ability that is best understood from several perspectives. In this chapter, we focus on neuropsychological investigations of ToM capacities that may be lateralized to one cerebral hemisphere; however, we locate these claims for lateralized components of ToM in an overview that includes many themes and that draws on several literatures. Our goal is to integrate ToM capacity into a general framework that incorporates philosophical, neuropsychological, and cognitive principles. We start with a brief statement of how ToM fits within current views on evolutionary psychology and developmental psychology to highlight a few critical features of ToM. Next, we provide background on localization and, in particular, lateralization of function. This selective review draws on work with monkeys as well as with humans. We emphasize the importance of the right hemisphere in ToM and related domains, but also discuss the role of bilateral prefrontal regions in ToM computations. Finally, and most importantly, we outline a neuropsychological model for ToM that incorporates contributions of prefrontal and limbic regions as well as the two hemispheres. The core ideas include the maintaince "decoupled" or alternative representations, selecting from among alternatives, and the relative salience of competing representations.

1. Evolution and the Social Brain

Proposals citing evolutionary pressure for social cognitive abilities have a long history (e.g., Chance & Mead, 1953; Jolly, 1966; and Kummer, 1968). In one influential paper, Humphrey (1976) portrayed nonhuman primate social networks as riddled with complex computational problems, far outweighing those found in predator-prey relations or those required for learning and reasoning about the physical world. Humphrey depicted primates as *homo psychologicus* or "natural psychologists," social tacticians who must take into account detailed knowledge of conspecifics who are likely opponents and possible allies. Survival and reproductive success seems likely to depend on being able to predict --reliably-- the behavior of conspecifics. Such complex problems, moreover, necessitate considerable neural resources, and thus offer a plausible evolutionary explanation for the huge brain-to-body ratio in the higher primates. Related work in evolutionary biology, such as Trivers' (1971) "reciprocal altruism" and Axelrod and Hamilton's (1981) "Tit-for Tat", proposed explanatory principles for survival strategies and behaviors in such networks. Dawkins and Krebs (1978; Krebs and Dawkins, 1984) suggested that communication should evolve to be manipulative, and to serve the genetic benefit of the sender. These theories began to merge into a unified picture. Byrne and Whiten (1988), with mounting evidence of tactical deception in primate social networks, coined the phrase "Machiavellian Intelligence" as an apt if not so flattering characterization of the evolution of human intelligence. In short, the extraordinary problem-solving capacities of the human brain appear to be fundamentally social in nature and, arguably, appear qualitatively similar in function to the problem solving abilities of related species. The similarity of social intelligence across different primates stands in contrast to the striking discontinuity between human and non-human generative language ability.

In addition to proposals about the evolutionary origins of ToM in humans, primate work has also stimulated methodological statements bearing on how best to characterize and test for ToM ability. When Premack and Woodruff (1978) first asked "Does the chimpanzee have a theory of mind?" they sought to show that our closest primate relatives predicted behavior in very much the same way we do; that is, by inferring the intentions and motivations of another. The extensive peer commentary on this paper, most notably by the philosophers Dennett (1978), Pylyshyn (1978), Bennett (1978), and Harman (1978), not only addressed a number of interesting epistemological and ontological issues regarding the Premack and Woodruff analysis, but put forth a number of suggestions carrying methodological weight as well. The now widely-accepted criterion for having a fully representational theory of mind, or "beliefs about beliefs," was set: predicting behavior based on the notion of a *false* belief.

2. The Child's Developing Ontology of Intentional States

The watershed task -- the prediction of behavior based on a *false* belief --is passed by children at roughly four years of age (e.g., Wimmer & Perner, 1983; Perner, Leekam, & Wimmer, 1987). This ability appears to emerge rather late when compared to children's everyday behaviors. Sullivan and Winner (1991), for example, relate the case of a two-year-old boy who, when asked by his grandmother *not* to touch the dials on the radio, immediately did so as soon as she left the room. Upon the grandmother's return the child said, "You didn't know this, but I was fixing it." While many utterances by young children concerning mental states may be dismissed as imitation or merely conversational (e.g., "know what?"), many are not so easily dismissed. Three-year-old children's spontaneous speech is riddled with talk of mental states *as* representations (Bartsch & Wellman, 1995). Yet, despite such ostensibly complex language, three-year-olds appear to be unable to appreciate and employ in the prediction of behavior, false-beliefs as causal entities.

Yet beliefs themselves, even false beliefs, are indeed items in three-year-olds' ontology, however fleeting and context-specific. Wellman and Bartsch (1988) have shown that three-year-olds can assign a causal role to beliefs if that belief does not conflict with what the child knows to be true. For example, if the child does not know the location of an object, or if an object is in two locations, only one of which was seen by the character, the child will then be able to predict where the character will look. It is important to note that there is no conflicting belief in this case as both are concordant with what the child believes to be real. Moreover, three-year old children who fail the false belief task are sometimes able to invoke false-beliefs as causal in their explanations (Bartsch and Wellman,1989).

A critical developmental milestone appears to be linked specifically to what can be termed *decoupled* representations, that is, the young child's ability to entertain simultaneous, alternative or conflicting beliefs. There are clear developmental precursors to a fully realized ability to manipulate alternative representations. In his developmental model of ToM, Leslie (1987; 1994; Leslie & Roth, 1993) posits a "decoupler" mechanism which matures between 18-24 months and facilitates the ability to represent representations; he calls this "metarepresentation" or "M-representation."[1] The first evidence of the M-representation system in action, according to Leslie, is observed in the pretense and shared pretense of young children. Shared pretense not only requires that the child hold in mind two contradictory representations concurrently (e.g., the banana is a telephone *and* the banana is a banana) but also that the child appreciate that a partner's behavior is caused by a fictional attitude (pretending) and relates to a fictional state of affairs (the banana is a phone). Because the child must disregard the literal meaning of a pretending speaker's utterance, and instead infer the speaker's intention, shared pretense may be considered an instance of Gricean communication (Grice, 1975; Leslie & Happé, 1989).[2] Based on such a formulation, the two-year old child could be said to have a theory of mind, despite not being able to pass the false belief task for another two years. (See also, Fodor, 1992; Baron-Cohen, 1995; Premack, 1990; and Whiten, 1996, for similar conclusions.)

Another key feature of the child's emerging ToM capabilities rests on the salience of the stimulus items used. Zaitchik (1991) has shown that three-year-old children are able to predict behavior based on a false belief if the salience of the stimulus is diminished. If children are not shown the location of an object, but merely told of its existence, they are able to make successful false belief predictions. The same has been shown for other, less salient, representations such as values (Flavell, Flavell, Green, & Moses, 1990) and intentions (Moses, 1990, as cited in Astington & Gopnik, 1991).

In sum, the prediction of behavior based on a false belief rests on at least two components: the ability to decouple representations and the ability to mark the salience of certain stimuli, either by inhibition or amplification. There is debate on when in the course of normal development these components first appear and when they are fully functional, but there is no disagreement involving their place in the adult ToM. There is likewise no disagreement that a person's proficiency using false beliefs to understand the behavior of others provides a good index of ToM ability. These components of ToM will figure prominently in our discussion.

3. Localization of Function within and across the Cerebral Hemispheres

Because of the gross anatomical symmetry of the two hemispheres, their functional differences received relatively little discussion until the latter half of the nineteenth century (Springer & Deutsch, 1998). The notion of hemispheric specialization is due in large part to the influence of Paul Broca. Broca's

---

[1] Leslie changed his terminology from meta- to M-representation after Perner (1991) took issue with his usage. Perner prefers to save the term "metarepresentation" to signify representations of representations *as* representations. The issue of what constitutes metarepresentation is currently the center of a lively and important debate. (See Sperber, forthcoming.)

[2] Genuine communication in the Gricean sense requires at least three orders of intentionality, e.g., I *intend* you to *recognize* that I *intend* you to do *x*. While this argument has been extensively worked out, it is a less parsimonious treatment of intentional attribution than is generally seen in developmental psychology (see also Bennett, 1976).

appreciation of the relationship between left hemisphere damage and speech abnormalities, as well as a correlation between handedness and speech, set the stage for the battles of cerebral dominance. Karl Wernicke's work on the role of the left temporal lobe in speech comprehension and Hugo Leipmann's studies in apraxia fueled the distinction and led to the proposal that the left hemisphere directed higher cognitive functions, such as language and purposeful movement, and played the "leading" role to the subdominant and unspecialized right hemisphere. For the next seventy years or so, until the early twentieth century, researchers concentrated on the localization of function within the left hemisphere, leaving the "minor" right largely ignored.

One obvious reason for this emphasis on the left hemisphere is the importance of language functions. Another, less obvious reason for this neglect may be due to the extent of damage needed for clinical symptoms to emerge; while even a small lesion to the left hemisphere may produce an aphasia or some other symptom, a similarly-sized lesion to the right hemisphere may not result in any detectable functional deficits (Semmes, 1968). There is as yet no definitive account for this difference between the hemispheres; however, there are several threads that can be woven together. It is now known that there are significant differences in the distribution of white and gray matter between the hemispheres. The appearance of "white" matter is due to glial sheaths that surround axons and obscure--to visual inspection--the blood in capillaries. Myelination of an axon increases the speed of transmission and is characteristic of neurons involved in long distance processing. The appearance of "gray" matter is due to the absence of myelination, that is, due to blood visible through the walls of capillaries. A higher proportion of gray matter is associated with relatively more cell bodies that send and receive information over short distances, that is, local processing. The left hemisphere is described as containing greater cell density, as having relatively more nonmyelinated fibers, especially in the frontal and precentral regions and as having more areas devoted to motor and specific sensory functions. The right hemisphere, in contrast, is described as having more areas of "associative" (higher level, integrative) cortex (Galaburda, 1995; Gur et al., 1980; Kertesz et al., 1990. See, Best, 1988, for a comprehensive account of cerebral asymmetries based on different growth gradients in neural development.)

These characteristics of the left hemisphere implicate its role in focused computations requiring faster, more localized processing involving collections of cells working together in close spatial proximity. Closely interconnected circuitry has the benefit of carrying out subtle alterations of a serial schematic action sequence, allowing for "many possible variations on any one cognitive theme" (Kinsbourne, 1982, p. 412). Moreover, proximal circuitry is ideal for fast, routinized processes such as syntactic analysis or speech production. Such small and quick variations are necessary for fine motor control performances such as speech production (e.g., the short delay in the onset of voicing between "pa" and "ba") and corresponding aspects of speech comprehension. Regarding comprehension, consider the difference between the consonant-vowel syllables "ba," "da," and "ga:" The primary distinctions among these are frequency changes occurring within the first 50 milliseconds. The left hemisphere has shown a clear advantage on such tasks (Kimura, 1967; 1993; Mattingly & Studdert-Kennedy, 1991). On this view, a lesion to the left hemisphere can affect a large proportion of the tissue employed in fast-acting, concerted processing and, as such, produce an obvious impairment.

In contrast, the right hemisphere possesses relatively more myelinated axons that, presumably, link different brain regions which, together, carry out functions that (1) do not require the same degree of localized, sequential processing or (2) require integration of different types of input initially processed in different parts of the brain. In this way, the integrative nature of the functions supported by the right hemisphere may require more extensive damage for clinical presentation in right hemisphere damaged (RHD) patients, at least insofar as concerns comprehension of propositional knowledge and language. Relatedly, patients with small, purely cortical right-sided strokes often recover quickly and well enough to resume their normal lives and, as a result, are not often seen by researchers.

4. Functions Associated with the Right Hemisphere

Before starting this selective review, we emphasize that the patients tested are, with few exceptions, stroke patients who have suffered unilateral brain damage somewhere in the territory of the middle cerebral artery. This area includes posterior portions of the frontal lobe, and portions of the temporal and parietal lobes as well. Localization within the right hemisphere is typically not discussed in detail in these papers. In what follows, we use the term "right hemisphere" as it has been used in the literature--to refer to regions anywhere within the hemisphere, though usually regions served by the middle cerebral artery and excluding the prefrontal regions.

History and Overview of Interpretations of Right Hemisphere Functions. The right hemisphere has been assigned a major role in some domains. The dominance of the right hemisphere in spatial reasoning became evident in the early part of this century, as RHD patients exhibited particular difficulty on visuospatial tasks. Similarly, RHD patients exhibited deficits in the production and perception of emotion. The left hemisphere came to be stereotyped as analytic, logical, local, and rational, while the right hemisphere as synthetic, Gestalt, holistic, global, and intuitive with respect to processing in several domains (Kolb & Whishaw, 1996). These characterizations, however, do not adequately capture the phenomenon of lateralization; the computations carried out within the right hemisphere require the same degree of precision as those in the left. Moreover, cerebral lateralization and specialization should not be viewed as a battle for dominance between the hemispheres, but more as a developmental streamlining of complementary functioning, much of which occurs in ontogeny and may not be specified in the genome (Deacon, 1997). Nevertheless, a sufficient number of regularities emerge to make lateralization of function a useful perspective for cognitive neuroscience.

The RHD literature on language impairment is wide-ranging and complex. Deficits include problems with integration of verbal as well as spatial material (Benowitz, Moya, & Levine, 1990), and also include verbosity, tangentiality, confabulation, concrete and fragmentary performance, difficulty with pragmatic inference, aberrant production and comprehension of humor, lack of self-monitoring in verbal responses, and difficulty in the interpretation and production of the prosodic and emotional dimensions of language (Beeman & Chiarello, 1998; Brownell & Gardner, 1988; Brownell & Martino, 1998; Brownell, Prather, Gardner, & Martino, 1995; Code, 1987; Gardner, Brownell, Wapner, & Michelow, 1983; Joanette, Goulet, & Hannequin, 1990; Myers, in press; Tompkins, 1994). We illustrate that ToM impairments fit easily within the range of deficits associated with right-sided brain injury. We will argue that a common feature of many reported deficits is that RHD patients exhibit difficulty processing decoupled or alternative interpretations of a stimulus, specifically with regard to the generation, synchronization, and marking of multiple representational sets.

The Role of the Right Hemisphere in ToM. Our first question is whether or not RHD affects a person's ToM. A set of recent studies (Brownell, Pincus, Blum, Rehak, & Winner, 1997; Happé, Brownell, & Winner, submitted; Winner, Brownell, Happé, Blum, & Pincus, 1998) establishes the nature of the deficit: RHD patients retain a basic ability to understand that people may harbor different beliefs about the world, but nonetheless have difficulty applying ToM to support comprehension.

A sample stimulus item used by Winner et al. (1998) serves to illustrate the deficit and how it has been tested with RHD patients and non brain-damaged control participants. The critical items all had the same structure built around a speaker's literally false statement concerning an awkward, potentially embarrassing situation ranging from cheating on a diet to cheating on one's spouse. The scenarios were designed to favor either of two interpretations: a deceitful lie uttered in self protection or an ironic joke uttered to lighten an awkward moment after one has been found out. In all scenarios, the addressee knows the truth of the situation. The critical distinction is that in joke scenarios, the speaker is *aware* of the addressee's knowledge while in lie scenarios the speaker is *unaware* of the addressee's knowledge. Questions posed during presentation of an item are presented below, along with their purpose. Additional factual questions were posed during presentation of the scenario to insure that participants maintained their focus on the task. If a participant answered the factual questions incorrectly, the vignette was presented again from the beginning. Items were presented to participants using a tape recorder and, simultaneously, in written form. A sample item is shown below:

> Sue smoked two packs of cigarettes a day. Zelda had been begging her to quit for years. Zelda promised that she would buy Sue dinner one night after work if Sue quit smoking for one week. Sue agreed to try.
>
> After two miserable days of not smoking, Sue began to sneak cigarettes in the bathroom. She told one friend, but she didn't tell Zelda.
>
> One day, Zelda came into the bathroom and saw smoke coming from one of the stalls. She recognized Sue's shoes and could tell that it was Sue who was in the stall smoking.
>
> QUESTION: Did Zelda realize that Sue was still smoking? Yes / No

This question measured a participant's sensitivity to a character's true first-order belief. Because all items included a true first-order belief, this question does not provide a sensitive index of first-order ability, but does guarantee that participants understand the content of a stimulus item.

> Zelda turned and left the bathroom before Sue could see her. Just then, Sue's friend, who was in the next stall, asked Sue, "Does Zelda know that you are still smoking?"
>
> QUESTION: What do you think Sue told her friend?
> a. "Yes, Zelda knows that I am still smoking."
> b. "No, Zelda does not know that I am still smoking."

This and the following question measure participants' understanding of the characters' second-order beliefs which could be true or false in different scenarios.

> QUESTION: Did Sue think that what she said was really true? Yes / No
>
> Sue returned to her desk. Zelda asked Sue, "Are you having trouble keeping your no-smoking promise?"
> Sue replied, "I haven't touched a cigarette all week."
>
> QUESTION: When Sue said that to Zelda, did she think that Zelda would believe her? Yes / No

This question measures a participant's ability to chart the implications of second-order beliefs.

> QUESTION: When Sue said, "I haven't touched a cigarette all week", was she:
> a. lying to avoid getting caught
> b. joking to cover up her embarrassment
>
> This question taps participant's ability to apply ToM to discourse interpretation, that Sue is lying.

In other comparable items, Zelda and Sue might confront one another in the women's room such that knowledge of the truth is clearly shared between the two:

> Sue opened the stall door with a lit cigarette in her hand and saw Zelda staring at her. Zelda turned and left the bathroom.... Sue returned to her desk. Zelda asked Sue, "Are you having trouble keeping your no-smoking promise?"
> Sue replied, "I haven't touched a cigarette all week."

Here, of course, Sue has been caught and must try to mend the effects of her behavior.

The most relevant results include a selective deficit in that the RHD patients were able to correctly answer fact questions but, on average, had more difficulty with discourse interpretation than the controls. Patients were less consistent in identifying second-order belief questions and less consistent in interpreting the final utterance as a joke or lie. In addition, the results confirm the link between second-order belief capacity and discourse interpretation. For both the RHD patients and the controls, a participant's degree of success distinguishing jokes from lies could be predicted very well (r = approximately +.7) from his or her facility with second-order beliefs.

Thus, second-order belief ability was shown to be reduced subsequent to RHD, and this reduction may well underlie these patients' poor discourse performance separating ironic jokes from lies. This study was designed specifically to test patients' sophistication with second-order belief and not first-order belief; first-order belief test questions *always* tapped true beliefs, which left unsettled the extent to which patients would also demonstrate impaired first-order belief deficits. Also left unsettled was the extent to which RHD patients' difficulties with ToM could be separated from general problems of inference.

The selectivity of the ToM deficit is supported by results from another study that examined the extent to which people consider the knowledge of an addressee as well as other features of a conversation. Brownell, Pincus, Blum, Rehak and Winner (1997) used terms of personal reference as a measure of discourse performance that is sensitive to what conversational partners know or do not know. Specifically, the task used in a pair of highly similar studies was to follow a conversation and, in that context, to choose an appropriate term to refer to someone not physically present. As illustrated below, in all stimulus conversations, the speaker and addressee were described as having equal, and relatively low, status, white collar jobs in which politeness would be the norm. The speaker and addressee were always described as having just met and having a conversation in the work environment about someone else, the referent. The referent was always male. In some items the referent was described as having an equal status (that is, the same occupation, for example, another clerk or museum guard) as the people talking, and in other items as having a clearly higher status (for example, a new vice president or director). The stimulus items also varied in terms of the knowledge possessed by the speaker, the addressee, or both were well acquainted with the referent. For example:

> You have just met Fran Hill, who, like you, is about to start working as a reservations clerk for American Airlines. You are describing your hopes about the job, when she says to you: "I heard that the other clerk who works in the reservations office is getting a big prize for having broken the reservations record this month." It then comes out in conversation between you and Fran Hill that neither of you has met Oliver Harding, the other clerk she is referring to.

The choices varied in terms of formality. We asked:

> Which of the following would you choose to respond to Fran Hill?
> (a) "Well, Oliver must be a really hard worker."
> (b) "Well, Mr. Harding must be a really hard worker."

A moment's reflection serves to identify the conventionally polite choices. In general, higher status people are referred to more formally. Beyond that static element of conversational context, an appropriate choice of a term of reference requires simultaneous consideration of both the speaker's and addressee's familiarity with the referent. Informality ("...Oliver...") is appropriate only in the one case in which both the speaker and referent are well acquainted with the referent. In all other cases, the more formal ("...Mr. Harding...") is canonically preferred. If only the speaker knows the referent, using the informal first name alone constitutes flaunting one's familiarity. If only the addressee knows the referent well, or if neither the speaker nor the addressee knows the referent, a speaker's use of the informal first name reflects presumptuousness. And, lastly, if both the speaker and addressee know the referent well, use of the formal title and last name seems a little odd, as if the speaker is trying to deny familiarity and to insert social distance from the referent.

Stimulus conversations were presented to RHD patients and to non brain-damaged control participants using a tape recorder and, simultaneously, on a typed script. In addition, simple outline drawings representing the referent, the speaker, and the addressee were kept in view to help patients maintain attention to the relevant characters. After a stimulus was presented, and before choosing a term of reference, the participant was asked factual questions about how well the speaker and listener knew the referent and about the status of the referent's job relative to the speaker and listener (that is, either higher or equal). Only after answering these factual questions correctly (which very rarely required repetition of the stimulus vignette) was the participant asked to choose an appropriate utterance to continue the conversation.

The dependent measure was a "formality score" defined as the proportion of times a participant selected the formal term of reference. Statistical analysis focused on two critical comparisons: (1) the consistency with which participants restricted their use of informal terms of reference to the one situation in which both the speaker and addressee knew the referent, and (2) the consistency with which participants used formal terms for high status referents.

The results were quite clear and equivalent across two highly similar studies. The RHD patients showed no deficit whatsoever in their use of status as a conversational variable. Both the patients and the nonbrain-damaged control participants consistently--and to the same degree-- used more formal terms to refer to high status people. However, the patients were far less consistent than the controls in incorporating knowledge states into their responses, even though in all cases the participants had answered the factual information correctly. (Specifically, there was a reliable group difference in contrast scores representing this critical comparison across the four combinations of speaker and addressee knowledge.) In these two studies, then, the RHD patients showed a deficit in restricting their use of informal terms to just the one, canonically appropriate context. Thus, the RHD patients exhibited a selective impairment in application of mental states that distinguished their decreased competence in ToM from their ability to incorporate other conversational elements into their responses.

Further documentation of the functional selectivity and--for the first time-- the anatomical specificity of ToM deficits is provided by a recent study by Happé, Winner, and Brownell (submitted;

1997). These authors tested the effects of right-sided damage on patients' understanding of first order false belief using verbal stories involving mental states and pictorial cartoons whose humor rested on mental states. Happé et al. were also able to test left-hemisphere damaged (LHD) aphasic patients using modified procedures on ToM tasks. And, most importantly, on one of their cartoon tasks, Happé et al. were able to use exactly the same procedure to test all groups (non brain-damaged controls, RHD patients, LHD patients) and were thus able to document the anatomical specificity of a ToM impairment.

The first task compared patients' comprehension of stories based in large part on ToM and, for comparison, control stories that required inference for understanding but did not require ToM. The items were those used in a functional imaging study carried out by Fletcher et al. (1995). A sample ToM story follows:

Example ToM Story

> A burglar who has just robbed a shop is making his getaway. As he is running home, a policeman on his beat sees him drop his glove. He doesn't know the man is a burglar; he just wants to tell him he dropped his glove. But when the policeman shouts out to the burglar, "Hey you! Stop!", the burglar turns around, sees the policeman and gives himself up. He puts his hands up and admits that he did the break-in at the local shop.

Test Question: Why did the burglar do that?

Example Control Story

> A burglar is about to break into a jewelers' shop. He skillfully picks the lock on the shop door. Carefully he crawls under the electronic detector beam. If he breaks this beam, it will set off the alarm. Quietly he opens the door of the store-room and sees the gems glittering. As he reaches out, however, he steps on something soft. He hears a screech and something small and furry runs past him, towards the shop door. Immediately the alarm sounds.

Test Question: Why did the alarm go off?

Participants were timed as they read each story silently. When finished, participants turned the page over and latency was recorded. Participants then answered the test question. Their responses were later scored for accuracy and completeness, which, in the case of the ToM items, included the success with which a response appropriately characterized a character's mental state.

The pattern of results across this and other tasks argues strongly for a selective ToM deficit in RHD patients. The control participants performed equivalently on the ToM stories and the Control stories, or they performed better on the ToM items. The RHD patients consistently performed worse (both in terms of latency to finish reading and in quality of explanation) on the ToM stories.

Happé et al. replicated and extended this finding using a cartoon task. Participants viewed a set of captionless cartoon pairs. Each pair consisted of an original cartoon and a cartoon identical in all respects but one: the humor element had been removed. The humor in half of the original cartoons was based on a person's ignorance or false belief. We refer to these as ToM cartoons. The humor in the other half was based on violation of either a physical or social norm. We refer to these as nonToM cartoons. The participant's task was to view each pair of items (one original and one altered, not funny version of the same cartoon) and to select the correct, funny item. Time to choose was recorded, as was the number of correct choices.

On this task, non brain-damaged controls, RHD patients, and LHD aphasic patients were tested. The LHD group was small (n=5) but homogeneous. All were diagnosed as having Broca's aphasia subsequent to lesions that included Broca's area and a variety of other structures in the left hemisphere. As is typical for patients with lasting Broca's aphasia, this group of Broca's patients had left frontal lesions that extended beyond Broca's area per se and that often included subcortical pathways. The overall pattern was clear and the same as observed for the stories. The non brain-damaged controls and the LHD patients performed equivalently on the two kinds of items while the RHD patients performed reliably worse on the ToM items.

A ToM deficit is the most parsimonious explanation for these patients' performances across different measures. The RHD patients performed consistently poorly on ToM items while the non brain-damaged controls and LHD patients showed no such selective deficit. Alternative accounts based on overall difficulty of inference demands are simply not consistent with these results. The findings obtained with the LHD Broca's patients, though preliminary due to the small sample size, argues for the anatomical specificity of ToM problems associated with RHD. Further corroboration comes from a recent study by Stone, Baron-Cohen, and Knight (in press) who report no ToM deficits in patients with unilateral left dorsolateral damage. Viewed in aggregate, this work argues for a consistent link between ToM deficits and right-sided damage.

Based on these results, it would be reasonable to expect increased right hemisphere activation during brain imaging studies of ToM tasks. However, the findings to date force a more complicated account. Of particular interest in this regard is the study by Fletcher et al., who used some of the same stimulus items as Happé et al (submitted; see also Goel, Grafman, Sadato, & Hallet, 1995). While heightened right-sided activity was measured during ToM tasks, especially in medial frontal regions (see Figure 3 and Table 4) and the anterior cingulate, increased activity was more consistently observed in the left medial prefrontal cortex (Brodmann's area 8 and 9) relative to a control condition. In the direct comparison between the ToM and Physical Story conditions, the Fletcher et al. study also indicates heightened right hemisphere activity in the posterior cingulate cortex and in the inferior right parietal cortex. Fletcher et al. did not interpret these as right hemisphere effects because the same areas did not show heightened activation in comparison to another control condition consisting of unrelated sentences, which undermined the conclusion that these right hemisphere areas had a specific link to ToM (p. 121). We agree with Fletcher et al.'s reasons for emphasizing those areas whose specific links to ToM were supported most convincingly by their data. We note, though, that their results do not provide a strong argument against right hemisphere involvement. Specifically, the demands of trying to integrate and thus remember the unrelated sentences may have required extra right hemisphere participation (Schacter, et al., 1996; Joanette et al., 1990). Given the subtraction methodology used in PET studies the extra task demands of the unrelated sentence condition could possibly explain the apparent inconsistency of right hemisphere involvement in ToM tasks. We are not surprised by the implications of left medial prefrontal involvement given the left hemsiphere's superiority in linguistic operations and the role of the prefrontal cortex in working memory, planning, and other so-called "executive" functions. (We discuss these functions in more detail below, in section 5, paying special attention to issues of laterality).

In other imaging work, Baron-Cohen, Ring, Moriarty, Schmitz, Costa and Ell (1994) found the right orbitofrontal cortex to be selectively activated when participants identified words having to do with the mind (e.g., think) as opposed to the body (e.g., hand). We are of course intrigued by this finding but are as yet unsure exactly how to fit this work in with the extant literature. Baron-Cohen et al carried out a region-of-interest hypothesis, examining only bilateral orbitofrontal and frontal polar regions but not the more posterior prefrontal regions; nor did the task have an inferential component. It is therefore difficult to directly compare this work to the Fletcher et al study or with the findings from the semantic accessing literature, which also focuses on the activity in more posterior prefrontal regions (Warburton et al., 1996).

Because the study of intentional attribution is relatively new to cognitive neuroscience, and there have only been a handful of lesion and imaging studies directly concerned with this topic, we should not expect a clear picture of this complex human ability to have emerged. Predicting the behavior of conspecifics or, for that matter, any intentional system (Dennett, 1987) is arguably our most exercised form of cognition, with novel problems occuring on a daily, even moment to moment, basis. Needless to say, the processing demands are multifarious and often quite extensive. The lack of direct evidence on this topic necessitates an appeal to the literature outside ToM, where regional processing biases are becoming clearer. Such an appeal may provide us with framework to predict, on both structural and functional dimensions, where and how ToM problems are being solved in the brain.

The Role of the Right Hemisphere in Domains Related to ToM. The ToM studies on RHD subjects can be viewed as an extension of a substantial body of work that provides both empirical support for the nature of the impairments and indicates how ToM fits into a larger view of cognition and communication that emphasizes the importance of alternative representations or interpretations. (See Beeman, 1998; Brownell, Gardner, Prather, & Martino, 1995; Burgess & Chiarello, 1996, for discussion.) Humor, for example, is an area of deficit for RHD patients. Verbal jokes often rest on reinterpretation of a punch line, that is, on appreciation of revised interpretations that reflect ToM. Consider the following joke that, in different forms. has served as a stimulus item in three studies (Bihrle, Brownell, Powelson, & Gardner, 1986; Brownell, Michel, Powelson, & Gardner, 1983; Shammi, 1997).

> The neighborhood borrower approached Mr. Smith one Sunday afternoon and said, "Say, Smith, are you using your lawnmower this afternoon?"
> "Yes, I am" Mr. Smith replied warily.
> *Punch line*: The neighborhood borrower replied, "Fine, then you won't be using your golf clubs. I'll just borrow them."

In Brownell et al. (1983), RHD patients were presented with the beginning of a joke and asked to select the correct, funny punch line among from alternatives that included non sequitur endings that were simply incongruent ("The grass is greener on the other side of the fence"), correct punchlines that were incongruent but that could be reinterpreted to fit with the joke, and straightforward continuations of the story that did not have any element of incongruity ("Gee, do you think I could borrow it when you're done?"). The major finding, which was replicated in later studies, was that RHD patients were prone to selecting the incongruous endings whether or not they could be reinterpreted to fit with a joke's beginning; RHD patients were not fooled by the straightforward endings but were drawn to the non sequitur endings. The patients apparently had an intact sense of the primary narrative requirement of verbal humor--that a punch line violates expectancies generated during the body of the joke. Yet, they had trouble processing the secondary meaning of a punch line that distinguished it from other candidate endings that were also incongruous. This study was designed and carried out without any consideration of ToM. However, the jokes used as the basis for stimuli all required reinterpretation of the punch line, and fully half of the items involved elements of mistaken beliefs, ignorance, or fooling a victim.

The Brownell et al. (1983) study included only RHD patients and nonbrain-damaged control participants. However, Shammi (1997) used the same items as part of a comprehensive study of humor in patients with different lesion sites. Her results underscored the anatomical specificity of the deficit: Right prefrontally damaged patients were impaired in punch line selection while non aphasic, left prefrontally damaged patients were not. Also, Bihrle et al. (1986) tested a group of aphasic patients with lesions distributed across the left hemisphere on a pictorial humor task and found that RHD but not LHD was associated with this inability to apprehend alternative meanings of humor. In addition, Bihrle et al. found that RHD patients were drawn to slapstick humor that did not require the same sophisticated appreciation of mental states or integration of disparate elements.

Our review suggests that ToM deficits contribute to RHD patients' problems comprehending some types of humor and, in addition, that ToM impairments represent one manifestation of a broader problem that extends beyond humor and beyond humor based on ToM. Brownell, Potter, Bihrle, and Gardner (1986), for example, reported that RHD patients had analogous impairments when adequate comprehension required revising interpretations of nonhumorous discourse. A typical, two-sentence, stimulus item follows.

> Sally became too bored to finish the history book.
> She had already spent five years writing it.

Most listeners generate an inference upon hearing the first sentence that Sally was reading a dull book. After the second sentence, however, normal listeners abandon their initial inference in order to incorporate both sentences under a single interpretation, that Sally is an author running out of enthusiasm for a long project. The RHD patients tested in this study seemed "normal" insofar as they generated the same initial inference as the controls, but the patients became stuck on the initial inference and often failed to abandon their initial interpretation in favor of an alternative that was, in the end, more consonant with both sentences. Comprehension was tested by presenting fact questions to insure that they had understood and retained the material in individual sentences and inference questions that tapped whether or not participants could revise their initial inferences. As suggested above, the RHD patients were distinguished from the control participants in their heightened tendency to respond *true* to test statements such as "Reading the history book bored Barbara" that were based on the first sentence in isolation, and in their decreased tendency to respond *true* to appropriate, unifying inferences such as "Barbara became bored writing a history book."

The deficit uncovered in this study was not simply one of an inability to integrate information across sentence boundaries. RHD patients were not distinguished from controls in their interpretation of sentence pairs in which the overall gist followed naturally from the order in which information was presented, as in the following example.

Johnny missed the wild pitch.
The windshield was shattered.

Here, patients and controls tended, appropriately, to endorse test sentences such as "A baseball broke the windshield" and to reject items such as "The car was in an automobile accident." (See Brownell, Carroll, Rehak, & Wingfield, 1992; Leonard, Waters, & Caplan, 1997, for other examples of RHD patients' successful integration of information across sentence boundaries.) The deficit can be characterized as trouble moving flexibly from one interpretation to another, a cognitive rigidity. More recent work by Tompkins and her colleagues (Tompkins, Baumgaertner, Lehman, & Fossett, 1997) suggests that RHD patients' comprehension deficits reflect an inability to construct alternative readings. Under either description of the deficit, RHD patients are limited by their decreased prowess with alternative interpretations, even when those interpretations do not rest clearly on mistaken beliefs, ignorance, or efforts to fool someone.

The characterization of RHD patients' deficit in terms of processing alternative meanings has also often been demonstrated using phrasal and lexical materials in a series of "off-line" (that is, unspeeded) tasks. For example, Winner and Gardner (1977; see also Myers & Linebaugh, 1981; VanLancker & Kempler, 1987, for similar studies) for example, have shown that RHD patients' comprehension of phrasal metaphors such as "he has a heavy heart" is surprisingly poor, even compared to that of LHD patients with aphasia. RHD patients tended to use literal interpretations and to have trouble acknowledging the nonliteral, alternative meaning of even these over learned phrases. Analogous results have been reported for single word polysemous stimuli such as "warm" which can refer literally to temperature or, more metaphorically, to one's personality (e.g., Brownell, Potter, Michelow, & Gardner, 1984; Brownell, Simpson, Potter, Bihrle, & Gardner, 1990). These early studies reinforce the asymmetry between the respective roles of the left and right hemispheres in processing stimuli defined in terms of having more than a single interpretation.

More recent work has used "on-line" experimental techniques to examine processing of different types of meaning relations by LHD and RHD brain-damaged patients and by the left and right hemispheres of neurolgically intact control participants. For example, Bottini et al. (1994) reported PET data corroborating the special role of the right hemisphere for processing of metaphoric alternative meaning in neurologically intact adults.

Another approach has been to use a variation on the semantic priming paradigm (e.g., Neely, 1991). The experimental task is to decide whether or not a string of letters (the "target") presented on a computer screen is a real word. A "prime" word is presented prior to the target. The prime may or may not be related to the target, and, more specifically, the prime can be related in different ways to the target. The interesting result is that participants respond more quickly to real word targets when they are preceded by a prime that is semantically related in specific ways. Recent work by several researchers has used lateralized field presentation in conjunction with priming paradigms with neurologically intact young adults to explore the different responses of the two hemispheres to semantic associations. In most of these studies, a prime word or words is presented at fixation, which results in the word being viewed in both the left visual field, which projects to the right hemisphere, and the right visual field, which projects to the left hemisphere. The target word is then presented to one side of the central fixation point, that is, to just one visual field, thereby isolating the contralateral hemisphere's processing of that target.

A large body of work provides solid support for differences in how the two hemispheres process meaning. (See Burgess & Chiarello, 1996; Burgess & Simpson, 1988; Chiarello, 1998, for reviews.) The right hemisphere is more likely than the left to process weak or diffuse associations and low frequency alternative meanings, and the right hemisphere maintains activation over longer prime--target intervals. The left hemisphere actively dampens or inhibits activation of alternatives and focuses on the contextually most dominant reading. Nakagawa (1991), for example, examined the effects of strong and weak associations (e.g., "pound" provides a strong association to "hammer", whereas "drop" provides a weak association to "hammer") on subsequent recognition of words presented to the right and left visual fields. Nakagawa found very different priming effects for each hemisphere. The left hemisphere performance benefited only from strong, frequent, associations, whereas even the weakest associations benefited right hemisphere performance. Beeman and his colleagues (Beeman, Friedman, Grafman, Perez, Diamond, & Lindsay, 1994; see Beeman, 1998, for a review) has reported analogous results described in terms of a right hemisphere predilection for "coarse grained" semantic coding. For example, when three distantly related primes ("foot", "cry", and "glass") are presented to both hemispheres using central presentation, target words ("cut") show more facilitation when presented to the right hemisphere than to the left hemisphere. In contrast, the left hemisphere shows greater response to a single direct prime ("scissors") than it did to a set of distantly related primes. The right hemisphere showed as much facilitation for a set of weakly associated primes as for a single direct prime.

The recurring theme in the work by several researchers is that the right hemisphere's contribution to semantic processing includes long lasting facilitation effects from even weakly associated primes. These same researchers have interpreted their work as providing an account for the discourse-level impairments of RHD patients and have in some cases performed parallel studies using brain damaged patients and non brain damaged controls (Beeman, 1993; 1998; Burgess and Chiarello, 1996; Burgess & Simpson, 1988; Chiarello, 1998). The left hemisphere excels at selecting and processing a single, dominant interpretation while inhibiting the others. In many linguistic and social contexts, the left hemisphere's focused approach works well. However, whenever there is not a single appropriate, highly activated interpretation, the right hemisphere will play a larger role: when several considerations must be integrated or when an initially attractive interpretation must be abandoned in favor of another.

5. Functions associated with Prefrontal Regions

The relevance of prefrontal regions to ToM is well established. Activation studies, such as reviewed in other chapters in this volume, provide strong support for prefrontal component of ToM. In addition, lesion studies document deficits in ToM tasks (see Stone, this volume). What bears closer examination is how functions associated with prefrontal regions relate to those attributed to the right hemisphere. It is not always possible to draw a clear distinction between the functions of the prefrontal regions and those of the right hemisphere. The catalogue of linguistic and cognitive impairments observed in RHD patients could be substituted, virtually without notice, into any review article on prefrontal impairments (McDonald, 1993). McDonald (1993) provides an important perspective on how this could be. One reason for this is that the literatures on RHD and frontal lobe disorders have remained largely independent of one another. Moreover, as one might suspect, the majority of RHD patients tested have most likely had damage to the frontal lobes. Studies on the prefrontal lobe have faced difficulties similar to those common to RHD studies. Prefrontal lesions, like RH lesions, are often difficult to notice, as they do not affect motor control or measures or standard measures of intelligence (Shallice & Burgess, 1991). A small lesion may have no noticeable affect on behavior. For classic behavioral pathology, such as uncontrolled and maladaptive behavior, apathy, perseveration, etc. (Luria, 1973), frontal damage may have to be quite extensive. It has been argued that many of these so-called classic symptoms may require global cerebral dysfunction (Canavan et al. 1985). Finally, it is often difficult to distinguish left- from right-hemispheric locus in the medial prefrontal areas, due largely to their proximity. Consequently, most treatments of prefrontal function do not emphasize lateralization of function. We will emphasize lateralization within the prefrontal regions when possible, although we acknowledge that these functional asymmetries are not as universally accepted as is the case for more posterior functions. In addition, we cite reports detailing the effects of well specified lesions in monkeys to provide clues to the functional differences and similarities of the various frontal regions in humans.

Tasks sensitive to prefrontal lesions involve tests of working memory, behavioral suppression, directed attention, and sensitivity to context. In monkeys, for example, lesions to the dorsolateral regions produce deficits in delayed response or delayed alternation tasks, where monkeys have difficulty inhibiting their tendency to search where they first found food, instead of searching where they just saw it hidden (Jacobsen, 1936). Likewise, on a task involving sampling food wells, Passingham (1985) found monkeys with dorsolateral lesions to perseverate, that is, returning to previously sampled wells while neglecting many unsampled ones. Monkeys with lesions to the ventromedial aspects of the prefrontal cortices are able to pass the above tasks, though they are unable to pass the delayed non-match to sample task which requires them to connect a reward (food) to the newer of two stimuli (Mishkin & Manning, 1978). Finally, posterior prefrontal lesions produce deficits in multipart tasks requiring conditional associations between two stimuli in order to find a reward. These tasks require the monkeys to reverse their expectations (e.g., if x then z; if y then not-z). Unlike the previous tasks, only the stimulus relationship of the reward --but not its location-- changed (Petrides, 1986). Prefrontal tasks specific to humans show the same patterns. For instance, the classic "executive function" task, the Wisconsin Card Sort, requires marking the relevance of previous stimuli or responses with regard to the current context (e..g, Kolb & Whishaw, 1996).

One noticeable characteristic of the above tasks is that they all involve marking the relevance of previously learned information to a novel context, and more specifically, they all involve a negative marking (e.g., not-X) of previous information. The ability to assign "not-X" to a representation in memory is necessary for shifting attention to alternative or opposite choices.[3] Deacon (1997) writes:

> One of the most salient common features of... tasks sensitive to prefrontal damage is that they all... involve shifting between alternatives or opposites, alternating place from trial to trial, shifting from one stimulus to a new one, or from one pairwise association to another depending on the presence of different cues... They all have to do with using information about something you've just done or seen *against itself*, so to speak, to inhibit the tendency to follow up that correlation and instead shift attention and direct action to alternative associations. Precisely because one association works in one context or trial, it is specifically excluded in the next trial or under different stimulus conditions (p. 263).

The appreciation of the role of the prefrontal cortices in attention, planning, inhibition, and working memory across domains has led researchers to deem these regions the brain's "senior executive" or locus of "executive function" (Goldman-Rakic, 1987; Shimamura, 1995; Joseph, 1996). The term "senior executive" is perhaps misleading in that it implies that there is a general decision-maker in the brain. It is important to note that the notion of a neural control structure need not be equated with decision making. Clark (1996) provides a useful softening of the notion of control structures, considering them "any neural substrate whose primary role is to modulate the activitity of other neural circuits, structures, or processes -- that is to say, any items or processes whose role is to control the inner economy rather than to track external states of affairs or to directly control bodily activity" (p. 136). The prefrontal cortex appears to be such a structure. Clark's treatment dispenses with the tacit homunculus in "executive" accounts, and it is the treatment we prefer.

With respect to lateralization of function, Milner et al. (1991) have shown that the right prefrontal cortex in humans is associated with memory for the temporal placement and frequency of events. Schacter et al. (1996) argue that right prefrontal regions are crucial for item-specific recognition as opposed to recognition based on the general characteristics of previously studied objects. Such activations may reflect effortful or intentional retrieval processes (see also Kapur et al., 1995). Other studies on memory retrieval have invited researchers to make similar left-right distinctions, such as "general event knowledge" versus "event-specific knowledge" (Conway & Rubin, 1993), "familiarity" vs. "recall or recollection" (Hintzman & Curran, 1994) or "gist" vs, "verbatim" (Brainerd et al., 1995). Consistent activation of the right frontal lobe during episodic retrieval tasks has been confirmed by recent PET studies (Tulving et al., 1994). Similarly, Deglin and Kinsbourne (1996), in a task requiring subjects to solve syllogisms, found that electroconvulsive suppression of the left hemisphere resulted in a marked sensitivity to the plausibility of the content of certain premises. Rather than focusing on the logical relations of the syllogism, subjects attempted to specify the details of the contents of the dubious propositions. Suppression of the right hemisphere did not produce the same reliance on personal experience or detailed knowledge of the propositions.

This work is consistent with our claims that the right hemisphere is necessary for the activation of extensive representational sets and, in this vein, for meaning that is implied or novel rather than explicit and based on existing routines. We should not expect a meaningful representation to be located in a single isolable substrate. Meaning, of course, consists in relationships. Implied meaning, moreover, requires a larger set of relationships or component parts than does manifest or literal meaning, thus increasing the demands of synchronization. We have already argued that the right hemisphere shows superiority in the realm of integrated or implied meaning, and the work of Schacter et al. (1996), among others, suggests the same is true for episodic and item-specific recognition memory.

The selection of an appropriate response to an ambiguous or multistep task requires the maintenance of heterogeneous representational sets for an extended period of time. The high degree of connectivity between the prefrontal regions to other, more posterior and subcortical brain regions, make

---

[3] The degree to which this marking neccessitates a strong, episodic-like representation to be held on-line for comparison is an open question. Depending on the task, the mechanism may require far fewer processing demands, such as a subtle weight change or attention shift. Thus resulting as a functional equivalent of marking "not-x" to the current context.

this region architectonically well-suited for the maintenance of the extensive network necessary for such computations (Goldman-Rakic, 1987; Pandya & Barnes, 1984) .

The multiple representations putatively maintained through the participation of association areas must be tagged or somehow made distinctive. Damasio (1989; 1994) has proposed such a mechanism regarding social decision making which he calls the "somatic marker" hypothesis: a circuit of neural mechanisms and substrates underlying the marking of values of past and potential behavioral schemas. Damasio suggests that the ventromedial and orbitofrontal cortices, with their extensive connections to the amygdala and various limbic regions, act as "convergence" zones, using the state of the soma (a combination of the state of viscera, internal milieu, and skeletal musculature) to mark the value of anticipated consequences and possible responses. A convergence zone contains a record (or, "set") of simultaneous activity throughout cortical and subcortical structures, the output of which goes to the amygdala and other various autonomic effectors, which in turn activate somatic states pertinent to the set. The distribution and signal intensity of a newly enacted state is perceived by the somatosensory cortices in conjunction with the original set. Patients with damage to ventromedial and orbital regions may lose the ability to use the state of the soma as a value marker for potential outcomes, resulting in the odd and often inappropriate social behavior observed in patients with prefrontal damage. Damasio further suggests that right frontal regions have an advantage in the maintenance of this biasing and winnowing network, specifically regarding the marking of punishing consequences.

There is substantial evidence that right (as opposed to left) frontal regions have a greater involvement in negative emotions such as linked to withdrawal behavior. Davidson, Gray, LeDoux, Levenson, Panksepp & Ekman(1994), Gray (1994), and others have proposed that the right frontal lobe is concerned with marking negative consequences while the left frontal regions guide approach-related behavior. EEG measures of resting anterior brain activity have found greater left-frontal activity to correlate with measures of approach-related behavior and greater right-frontal activity to correlate with withdrawal (Sutton & Davidson, 1997). Similarly, in laboratory studies designed to manipulate emotion through various means (e.g., films, monetary reward, and punishment, etc.), pleasant states were associated with greater left activation while unpleasant states showed increased right-sided anterior activation. Similar results have obtained in nonhuman primate studies, where abnormally fearful monkeys show greater EEG activation in the right frontal lobes (Kalin, Shelton, Rickman & Davidson, 1998). Davidson speculates that, regarding the prefrontal modulation of the amygdala and other limbic regions, the left prefrontal cortex has an advantage in shutting down negative stimuli more quickly.

Damage to these regions may present with pathological amplification of this distinction: right frontal lesions are associated with manic-like excitement, with confabulatory and prosodically abnormal speech, while left frontal lesions are more often associated with apathy, depression, schizophrenic-like behaviors, and, especially with damage in medial areas, reduced speech, mutism, and catatonia (Robinson & Downhill, 1995; Gainotti, 1972). These symptoms suggest an amplification of the characteristics of the non-damaged hemispheres.

Interestingly, as suggested by Kinsbourne (1982), the affectively charged approach-withdrawal distinction may relate to the lateralization of sensorimotor function and also to lateralization of higher cognitive functions including focal attention. To appreciate our interpretation of Kinsbourne's model, it is important to note that all vertebrates, and motile organisms in general, possess lateralized sensorimotor functions that support an essentially free range of motion during approach and withdrawal behaviors. The successful detection of predator or prey requires bilateral peripheral pre-attentive monitoring. Even simple organisms, whose behavior is largely driven by external stimuli, must be able to represent an object in a specific location in space, which in humans is typically associated with right parietal regions. However, simple organisms may lack the cognitive resources to bring attention, anticipation, and many features of past experiences to bear on decision making processes, which in humans are associated with the prefrontal regions and, perhaps, with left more than right prefrontal regions. Kinsbourne writes:

> ...behaviorally more elaborately equipped species can cost account their decision making more exactly. Having oriented to a stimulus perceived in meager detail in the peripheral sensory field, they bring focal attention to bear to make finer distinctions before committing themselves to action. As the object is scrutinized. . . two processes proceed in parallel: Information is extracted serially, feature by feature; and concurrently, the location of each feature is registered on a centrally represented spatial framework, relative to the feature locations already represented (p.412).

Kinsbourne acknowledges that information processing with regard to feature extraction or directed focal attention is typically mediated by the left hemisphere, while relational information is typically lateralized to the right. (See also work by Arguin, Joanette, & Cavanagh, 1993; and Treisman & Gelade, 1980.) Once focal attention begins, such as in approach behaviors, the need to maintain relational information between the organism and the object is relaxed because the need to effectively spatially orient to the object, by this point, is past. Likewise, withdrawal behaviors do not necessitate, or even allow for, the same kind of serial feature extraction processing as approach behaviors. During avoidance behaviors, information about the relation between the organism and the object must be maintained over time and constantly updated for successful withdrawal to occur. Thus, we are drawn to the notion that the lateralized roles of prefrontal and posterior regions we have discussed in terms of symbolic and emotional processing have quite simple and fundamental phylogenetic roots.

6. Outline of a Neuropsychological Account of Theory of Mind

The neuropsychological literature on ToM has focused on two large regions of the human brain, the prefrontal cortex and the right hemisphere. Both areas have been implicated in reports on working memory, executive function, comprehension of discourse including humor, insight, interpersonal skills, as well as ToM (e.g., McDonald, 1993; Alexander et al., 1989). We believe that a large portion of the observations in the literature can be incorporated into a framework that includes, but is not limited to, ToM. We note also that many of the functional features that we have attributed to right hemiphere and prefrontal regions appear to be absent or diminished in autistic patients. The Wisconsin Card Sort, Tower of Hanoi, and other classic tests of prefrontal function have proven difficult for autistic patients, even for those who are able to pass first- and second-order false belief tasks (Ozonoff, Pennington, & Rogers, 1991; Russell, Mauthner, Sharpe, & Tidswell, 1991; Happé, this volume). Difficulty with the interpretation of non-literal speech is also symptomatic of autism and has been implicated as playing a role in their ToM deficits (Mitchell and Isaacs, 1994). Language pragmatic factors have also been suggested to underlie normal children's and RHD patients' ToM deficits (Siegal, Carrington, & Radel, 1996).

The prefrontal regions are critical for ToM in that they support a person's ability to select from among representations that are divorced from sensory input. The literature identifies the prefrontal regions as relevant to working memory, aspects of attention, marking for salience, and selection. Social decision making, including ToM, rests on conflicting information that must be maintained and evaluated; the most relevant information must be highlighted, and the less relevant information must be inhibited for successful interpretation to occur. The architectonic changes in the human brain over the course of recent evolution are such that the prefrontal cortex, with its extensive connections with both cortical and subcortical regions, is well suited for this role. These structural and functional attributes are entirely consistent with work by Damasio (1994) as well as others, and consistently point to the prefrontal regions' contributions to multi-step decision making processes, and how the right prefrontal regions may have a larger role for "hot" or more affectively-laden ToM problems. (See Brothers & Ring, 1992.) The many connections to and from the prefrontal areas make these regions the obvious choice for modulating information from different sources and modalities, and implicate their role in the winnowing and amplification of possible behavioral options.

The second core idea is that the right cerebral hemisphere is vital for humans' ability to maintain multiple representational sets. ToM, and competence in a host of other cognitive domains, depends crucially on a person's ability to refer to alternative, often conflicting, interpretations of the same topic. While the left hemisphere inhibits or dampens alternatives, the right hemisphere, we argue, is critical for people's ability to maintain access to alternative representations including those involved in figurative language, jokes, and stories as well as mental states. They may be weak semantic associates as examined by Beeman (1998), alternative readings of a punch line as examined by Shammi (1997) or Bihrle et al. (1986), or different mental states as examined by Happé et al. (submitted) or Winner et al. (1998).

Taken together, evidence from the many literatures addressed above presents a compelling case for lateralized information-processing biases. Much of the time, the mechanisms underlying our ToM, or our "folk psychology" in general, appear to operate effortlessly; we simply "know" what another is thinking, how he or she is feeling, and consequently, what that person is about to do. At other times, however, the process is not so automatic; we belabor possible scenarios, perhaps wondering what we would do if faced with the same situation, or perhaps relying on our past experience with that person or with people in general. As Dennett (1987; 1998) has repeatedly stressed, our ability to predict behavior sucessfully through ToM (what he calls the "intentional stance") relies on the *rationality* of behavior. Our well-crafted tools of mental state attribution would be rendered useless if behavior did not, for the most part, fit our folk psychological stereotypes. Yet daily life presents us with a variety of social scenarios, many stereotypical, even script-like, and many of which are novel, and quite complicated indeed. Thus the neural substrates underlying the process of inferring the content of mental states, on our view, will vary depending on the demands of the task at hand. Novel tasks that require maintaining and updating relational and ambiguous information will necessitate right-hemisphere involvement, while requirements of the more routine, script-like ToM tasks may be processed largely by left hemisphere regions. Moreover, as suggested by Kinsbourne (1982), Deglin and Kinsbourne (1996), and Arguin et al. (1993), we can expect more left activation for tasks requiring focal attention or non-conflicting syllogistic reasoning.

As detailed in other chapters in this volume, research on the neural substrates underlying ToM has been informed by studies with lesion populations, abnormal phenotypes, and various in vivo imaging techniques. It is important to note that what ties these paradigms together in the present context is a functional analysis, which is extrinsic to the anatomy. The same problems may be solved using different neural mechanisms, even in the same brain, and a complex problem will have many component parts, each of which may implicate a different brain region or structure. Nonetheless, converging research from many areas has contributed to our understanding of the working brain, and the heterogeneous processes that underlie our folk psychology.

In sum, though the literatures on RHD, frontal pathology, and activation in intact brains are at times difficult to interpret, some reliable generalizations emerge. We argue that the right hemisphere and prefrontal regions both are required for normal ToM; impairment in either can disrupt task performance. The importance of the right hemisphere's contribution to these aspects of social cognition seems to depend on several factors. The need to maintain alternative readings or tangential associations for longer periods of time, the novelty of the situation (that is, the absence of an appropriate decision making algorithm), and the affective marking of an alternative are all features that will increase the potential contribution of the right hemisphere.

We note in conclusion some implications that invite empirical test. It should be possible to amplify right hemisphere contributions to ToM by using scenarios which require affective marking by the participant. Similarly, right hemisphere involvement should be heightened when an inference about a person's mental state requires integrating conflicting or ambiguous information, or information that requires detailed episodic memory. In contrast, right hemisphere involvement should be minimized when mental states can be inferred on the basis of highly salient, non contradictory information. In general, the right hemisphere helps preserve the raw material from which the prefrontal and limbic regions draw in decision making processes. Theory of mind thus provides a rich framework for understanding social cognition from a neuroscientific perspective.

References

Alexander, M. P., Benson, F. D., & Stuss, D. T. (1989). Frontal lobes and language. *Brain and Language, 37*, 656-691.

Arguin, M. Joanette, Y., & Cavanagh, P. (1993). Visual search for feature and conjunction targets with an attention deficit. *Journal of Cognitive Neuroscience, 5*, 436-452.

Astington, J. W., & Gopnik, A. (1991). Theoretical explanations of children's understanding of the mind. *British Journal of Developmental Psychology, 9,* 7-31.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211*, 1390-1396.

Baron-Cohen, S. (1995). *Mindblindness*. Cambridge: MIT Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*, 37-46.

Bartsch, K., & Wellman, H. M. (1989). Young children's attribution of action to beliefs and desires. *Child Development, 60*, 946-964.

Bartsch, K., & Wellman, H. M. (1995). *Children talk about the mind*. New York: Oxford University Press.

Beeman, M. (1993). Semantic processing in the right hemsiphere may contribute to drawing inferences during comprehension. *Brain and Language, 44*, 80-120.

Beeman, M. (1998). Coarse semantic coding and discourse comprehension. In M. Beeman & C. Chiarello (Eds.), *Right hemisphere language comprehension: perspectives from cognitive neuroscience* . Mahwah, NJ.: Lawrence Erlbaum Associates.

Beeman, M., & Chiarello, C. (1998). *Right hemisphere language comprehension: Perspectives from cognitive neuroscience*. Mahwah, NJ: Lawrence Erlbaum.

Beeman, M., Friedman, R. B., Grafman, J., & Perez, E. (1994). Summation priming and coarse semantic coding in the right hemisphere. *Journal of Cognitive Neuroscience, 6*, 26-45.

Bennett, J. (1976). *Linguistic Behavior*. Cambridge: Cambridge University Press.

Bennett, J. (1978). Some remarks about concepts. *The Behavioral and Brain Sciences, 1*, 557-560.

Benowitz, L. I., Moya, K. L., & Levine, D. N. (1990). Impaired verbal reasoning and constructional apraxia in subjects with right hemisphere damage. *Neuropsychologia, 28*, 231-241.

Best, C. T. (1988). The emergence of cerebral asymmetries in early human development: A literature review and a neuroembryological model. In D. L. Molfese & S. J. Segalowitz (Eds.), *Brain lateralization in children: Developmental implications* (pp. 5-34). New York: Guilford Press.

Bihrle, A. M., Brownell, H. H., Powelson, J. A., & Gardner, H. (1986). Comprehension of humorous and non-humorous materials by left and right brain damaged patients. *Brain and Cognition, 5*, 399-412.

Bottini, G., Corcoran, R., Sterzi, R., Paulesu, E., Schenone, P., Scarpa, P., Frackowiak, R. S., & Frith, C. D. (1994). The role of the right hemisphere in the interpretation of figurative aspects of language: a positron emission tomography activation study. *Brain, 117*, 1241-1253.

Brainerd, C. J., Reyna, V. F., & Kneer, R. (1995). False-recognition reversal: When similarity is distinctive. *Journal of Memory and Language, 34*, 157-185.

Brothers, L., & Ring, B. (1992). A neuroethological framework for the representation of minds. *Journal of Cognitive Neuroscience, 4*(2), 107-118.

Brownell, H. H., Carroll, J. J., Rehak, A., & Wingfield, A. (1992). The use of pronound anaphora and speaker mood in the interpretation of conversational utterances by right hemisphere brain-damaged patients. *Brain and Language, 43*, 121-147.

Brownell, H. H., & Gardner, H. (1988). Neuropsychological insights into humour. In J. Durant, J. Miller (Eds.), *Laughing matters: A serious look at humour* (pp. 17-34). London: Longman Technical.

Brownell, H., Gardner, H., Prather, P., & Martino, G. (1995). Language, communication, and the right hemisphere. In H. S. Kirshner & et al. (Eds.), *Handbook of neurological speech and language disorders* (pp. 325-349). New York, NY, USA: Marcel Dekker, Inc.

Brownell, H. H., & Martino, G. (1998). Deficits in inference and social cognition: The effects of right hemisphere brain damage on discourse. In M. Beeman & C. Chiarello (Eds.*), Right hemisphere language comprehension: Perspectives from cognitive neuroscience* (pp. 309-328). Mahwah, NJ: Erlbaum.

Brownell, H. H., Michel, D., Powelson, J., & Gardner, H. (1983). Surprise but not coherence: sensitivity to verbal humor in right- hemisphere patients. *Brain and Language, 18*, 20-27.

Brownell, H., Pincus, D., Blum, A., Rehak, A., & Winner, E. (1997). The effects of right-hemisphere brain damage on patients' use of terms of personal reference. *Brain and Language, 57*, 60-79.

Brownell, H., Potter, H. H., Birhle, A. M., & Gardner, H. (1986). Inference deficits in right brain damaged patients. *Brain and Language, 27*, 310-321.

Brownell, H. H., Potter, H. H., Michelow, D., & Gardner, H. (1984). Sensitivity to lexical denotation and connotation in brain-damaged patients: A double dissociation? *Brain and Language, 22*, 253-265.

Brownell, H. H., Simpson, T. L., Bihrle, A. M., Potter, H. H., & Gardner, H. (1990). Appreciation of metaphoric alternative word meanings by left and right brain-damaged patients. *Neuropsychologia, 28*, 375-383.

Burgess, C., & Chiarello, C. (1996). Neurocognitive mechanisms underlying metaphor comprehension and other figurative language. *Metaphor and Symbolic Activity, 111*, 67-84.

Burgess, C., & Simpson, G. B. (1988). Cerebral hemispheric mechanisms in the retrieval of ambiguous word meanings. *Brain and Language, 33*, 86-103.

Byrne, R., & Whiten, A. (1988). The Machiavellian intelligence hypotheses: editorial. In R. Byrne & A. Whiten (Eds.), *Machiavellian Intelligence: Social expertise and the evolution of intellect in monkeys, apes, and humans* . Oxford: Clarendon Press.

Canavan, A., Janota, I., & Schura, P. H. (1985). Luria's frontal lobe syndrome: psychological and anatomical considerations. *Journal of Neurology, Neurosurgery and Psychiatry, 48*, 1049-1053.

Chance, M. R. A., & Mead, A. P. (1953). Social behavior and primate evolution. *Symposia of the Society for Experimental Biology, 7*, 395-439.

Chiarello, C. (1998). On codes of meaning and the meaning of codes: Semantic access and retrieval within and between hemispheres. In M. Beeman & C. Chiarello (Eds.), *Right hemsiphere language comprehension: perspectives from cognitive neuroscience* (pp. 141-160). Mahwah, NJ.: Lawrence Erlbaum Associates.

Clark, A. (1996) *Being there: putting brain, body and world back together again*. Cambridge, MA: MIT Press.

Code, C. (1987). *Language, aphasia, and the right hemisphere*. New York: Wiley.

Conway, M. A., & Rubin, D. C. (1993). The structure of autobiographical memory. In A. F. Collins, S. E. Gathercole, M. A. Conway, & P. E. Morris (Eds.), *Theories of memory* . Hillsdale, N.J.: Lawrence Erlbaum.

Damasio, A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation, 1*, 123-132.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Grosset/Putnam.

Davidson, R. J. (1992). Anterior cerebral asymmetry and the nature of emotion. *Brain and Cognition, 20*, 125-151.

Davidson, R., Gray, J., LeDoux, J., Levenson, R., Panksepp, J., & Ekman, P. (1994) Is there emotion-specific physiology? In P. Ekman & R, Davidson (Eds.) *The nature of emotion: Fundamental questions* (pp. 235-262). New York: Oxford University Press.

Dawkins, R., & Krebs, J. R. (1978). Animal signals: information or manipulation? In J. R. Krebs & N. B. Davies (Eds.), *Behavioral ecology: An evolutionary approach* (pp. 282-314). Oxford: Blackwell.

Deacon, T. W. (1997). *The symbolic species: the co-evolution of language and the brain*. New York: W.W. Norton & Company.

Deglin, V. L., & Kinsbourne, M. (1996) Divergent thinking styles of the hemispheres: how syllogisms are solved during transitory hemiphere suppression. *Brain and Cognition, 31*, 285-307.

Dennett, D. C. (1978). Beliefs about beliefs. *The Behavioral and Brain Sciences, 1*, 568-570.

Dennett, D. C. (1987*). The intentional stance*. Cambridge, MA.: MIT Press.

Dennett, D. C. (1998). *Brainchildren: Essays on designing minds*. Cambridge, MA.: MIT Press.

Flavell, J. H., Flavell, E. R., Green, F. L., & Moses, L. J. (1990). Young children's understanding of fact beliefs versus value beliefs. *Child Development, 61*, 915-928.

Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R., Frackowiak, R. S. J., & Frith, C. (1995). Other minds and the brain: a functional imaging study of theory of mind in story comprehension. *Cognition, 57*, 109-128.

Fodor, J. (1992). A theory of the child's theory of mind. *Cognition, 44*, 283-296.

Gainotti, G. (1972). Emotional behavior and hemisphere side of lesion. *Cortex, 8*, 41-55.

Galaburda, A. (1995). Anatomic basis of cerebral dominance. In R. J. Davidson & K. Hughdahl (Eds.*), Brain asymmetry* (pp. 51-73). Cambridge, MA.: MIT Press.

Gardner, H., Brownell, H., Wapner, W., & Michelow, D. (1983). Missing the point: The role of the right hemisphere in the processing of complex linguistic materials. In E. Perecman (Ed.), *Cognitive processing in the right hemisphere* (pp. 169-191). New York: Academic Press.

Goel, V., Grafman, J., Sadato, N., & Hallett, M. (1995) Modeling other minds. *NeuroReport, 6*, 1741-1746.

Goldman-Rakic, P. S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In F. Plum (Ed.), *Handbook of Physiology: The Nervous System* (Vol. 5, pp. 373-417). Bethesda, Md.: American Physiological Society.

Gray, J. A. (1994). Three fundamental emotion systems. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 243-247). New York: Oxford University Press.

Grice, H. P. (1975). Logic and converation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol 3. Speech acts* (pp. 41-58). New York: Seminar Press.

Gur, I. K., Packer, J. P., Hungerbuhler, J. P., Reivich, M., Orbist, W. D., Amarnek, W. S., & Sackheim, H. A. (1980). Differences in the distibution of gray and white matter in human cerebral hemispheres. *Science, 207*, 1226-1228.

Happé, F., Brownell, H., & Winner, E. (submitted). Acquired theory of mind impairments following right hemisphere stroke.

Harman, G. (1978). Studying the chimpanzee's theory of mind. *The Behavioral and Brain Sciences, 1*, 591.

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgements: evidence for separate processes of familiarity and recall. *Journal of Memory and Language, 33*, 1-18.

Humphrey, N. (1976). The social function of the intellect. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing points in ethology* . Cambridge: Cambridge University Press.

Jacobsen, C. (1936). Studies of cerebral functions in primates. *Comparative Psychology Monographs, 13*, 1-68.

Joanette, Y., Goulet, P., & Hannequin, D. (1990). *Right hemisphere and verbal communication.* New York: Springer-Verlag.

Jolly, A. (1966). Lemur social behaviour and primate intelligence. *Science, 153*, 501-506.

Joseph, R. (1996). *Neuropsychiatry, neuropsychology, and clinical neuroscience: Emotion, evolution, cognition, language, memory, brain damage, and abnormal behavior (2nd ed.).* Baltimore, MD, USA: Williams and Wilkins Co.

Kalin, N. H., Shelton, S. E., Rickman, M., & Davidson, R. (1998). Individual Differences in Freezing and Cortisol in Infant and Mother Rhesus Monkey. *Behavioral Neuroscience, 112*, 251-254.

Kapur, N., Friston, K. J., Young, A., & Frith, C. D. (1995) Activation of human hippocampal formation during memory of faces: A PET study. *Cortex, 31*, 99-108.

Kertesz, A., Polk, M., Black, S. E., & Howell, J. A. (1990). Sex, handedness and the morphometry of cerebral asymmetries on magnetic resonance imaging. *Brain Research, 530*, 40-48.

Kimura, D. (1967). Functional asymmetry of the brain in dichotic listening. *Cortex, 3*, 163-168.

Kimura, D. (1993). *Neuromotor mechanisms in human communication.* New York: Oxford University Press.

Kinsbourne, M. (1982). Hemispheric specialization and the growth of human understanding. *American Psychologist, 37*, 411-420.

Kolb, B., & Wishaw, I. Q. (1996) *Fundamentals of human neuropsychology (4th ed).* New York; W. H. Freeman.

Krebs, J. R., & Dawkins, R. (1984). Animal signals: mind reading and manipulation. In J. R. Krebs & R. Dawkins (Eds.), *Behavioural ecology: an evolutionary approach* (pp. 380-401). Blackwell.

Kummer, H. (1968). *Social organization of hamadryas baboons*. Chicago: University of Chicago Press.

Leonard, C. L., Waters, G. S., & Caplan, D. (1997). The use of contextual information by right brain-damaged individuals in the resolution of ambiguous pronouns. *Brain and Language, 57*(3), 309-342.

Leslie, A. (1994). ToMM, ToBy, and Agency: Core architecture and domain specificity. In L. Hirschfied & S. Gelman (Eds.), *Mapping the mind: domain specificity in cognition and culture* (pp. 119-148). New York: Cambridge University Press.

Leslie, A. M. (1987). Pretense and representation: the origins of 'theory of mind'. *Psychological Review, 94*, 412-426.

Leslie, A. M., & Happé, F. (1989). Autism and ostensive communication: The relevance of metarepresentation. *Development and Psychopathology, 1*, 205-212.

Leslie, A. M., & Roth, D. (1993). What autism teaches us about metarepresentation. In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds.), *Understanding other minds: Perspectives from autism* . New York: Oxford University Press.

Luria, A. R. (1973). The frontal lobes and the regulation of behavior. In K. H. Pribram & A. R. Luria (Eds.), *Psychophysiology of the Frontal Lobes* (pp. 3-26). New York: Academic Press.

Mattingly, I., & Studdert-Kennedy, M. (1991). *Modularity and the Motor Theory of Speech Perception*. Hillsdale, N.J.: Erlbaum.

McDonald, S. (1993). Viewing the brain sideways? Frontal versus right hemisphere explanations of non-aphasic language disorders. *Aphasiology, 7*(6), 535-549.

Milner, B., Corsi, P., & Leonard, G. (1991). Frontal lobe contributions to recency judgements. *Neuropsychology, 29*, 601-618.

Mishkin, M., & Manning, F. (1978). Nonspatial memory after selective prefrontal lesions in monkeys. *Brain Research, 143*, 313-323.

Mitchell, P. & Isaacs, J. E. (1994) Understanding of verbal representation in children with autism: the case of referential opacity. *British Journal of Developmental Psychology, 12*, 439-454.

Moses, L. J. (1990). Young children's understanding of intention and belief. Unpublished doctoral dissertation. Stanford University.

Myers, P. S. (in press). *Right hemisphere damage: Disorders of communication and cognition.* San Diego: Singular Publishing Group.

Myers, P. S., & Linebaugh, C. W. (1981). Comprehension of idiomatic expressions by right-hemisphere damaged adults. In R. H. Brookshire (Ed.), *Clinical aphasiology: conference proceedings* (pp. 254-261). Minneapolis: BRK Publishers.

Nakagawa, A. (1991). Role of anterior and posterior attention networks in hemispheric asymmetries during lexical decisions. *Journal of Cognitive Neuroscience, 3*, 315-321.

Neely, J. H. (1991). Semantic Priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264-336). Hillsdale, NJ.: Lawrence Erlbaum.

Ozonoff, S., Pennington, B. F., & Rogers (1991) Executive function deficits in high-functioning autistic individuals: relationship to theory of mind. *Journal of Child Psychology and Psychiatry, 32*, 1081-105.

Pandya, D., & Barnes, C. (1987). Architecture and connections of the frontal lobe. In E. Perecman (Ed.), *The Frontal Lobes Revisted* (pp. 41-72). New York: IRBN Press.

Passingham, R. E. (1985). Memory of monkeys (*Macaca mulatta*) with lesions in prefrontal cortex. *Behavioral Neuroscience, 99*, 3-21.

Perner, J. (1991). *Understanding the representational mind.* Cambridge, MA: MIT Press.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief. *British Journal of Developmental Psychology, 5*, 125-137.

Petrides, M. (1986). The effect of periarcuate lesions in the monkey on the performance of symmetrically and asymmetrically reinforced visual and auditory go, no-go tasks. *Journal of Neuroscience, 6*, 2054-2063.

Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition, 36*, 1-16.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences, 1*, 515-526.

Pylyshyn, Z. W. (1978). When is attribution of beliefs justified? *The Behavioral and Brain Sciences, 1*, 592-593.

Robinson, R. G., & Downhill, J. F. (1995). Lateralization of psychopathology in response to focal brain injury. In R. J. Davidson & K. Hugdahl (Eds.) *Brain asymmetry* (pp. 693-711). Cambridge, MA.: MIT Press.

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The 'windows task' as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology, 9*, 331-50

Sackheim, H., Greenberg, M. S., Weiman, A. L., Gur, R. C., Hungerbuhler, J. P., & Geshwind, N. (1982). Hemispheric asymmetry in the expression of positive and negative emotions: Neurological evidence. *Archives of Neurology, 39*, 210-218.

Schacter, D. L., Curran, T., Galluccio, L., Milberg, W. P., & Bates, J. (1996). False recognition and the right frontal lobe: A case study. *Neurospsycholgia, 34*, 793-808.

Semmes, J. (1968). Hemispheric specialization: a possible clue to mechanism. *Neuropsycholgia, 6*, 11-26.

Shallice, T., & Burgess, P. (1991). Higher-order cognitive impairments and frontal lobe lesions in man. In H. S. Levin, H. M. Eisenberg, & A. L. Benton (Eds.), *Frontal lobe function and dysfunction* (pp. 125-138). New York: Oxford University Press.

Shimamura, A. P. (1995). Memory and frontal lobe function. In M. Gazzaniga (Ed.), *The cognitive neurosciences* . Cambridge, MA: MIT Press.

Shammi, P. (1997). Humor and brain damage. Unpublished doctoral dissertation. University of Toronto, Toronto, Canada.

Siegal, M., Carrington, J. & Radel, M. (1996). Theory of mind and pragmatic understanding following right hemisphere damage. *Brain and Language, 53*, 40-50.

Sperber, D. (Ed.) (forthcoming*). Metarepresentation: proceedings of the tenth cognitive science society.* New York: Oxford University Press.

Springer, S., & Deutsch, G. (1998). *Left brain, right brain: Perspectives from cognitive neuroscience (5th ed).* New York: W.H. Freeman and Company.

Stone, V., Baron-Cohen, S., Knight, R. (in press) Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience.*

Sullivan, K., & Winner, E. (1991). When 3-year-olds undertand ignorance, false belief and representational change. *British Journal of Developmental Psychology, 9.*

Sutton, S. K., & Davidson, R. J. (1997). Prefrontal brain asymmetry: a biological substrate of the behavioral approach and inhibition systems. *Psychological Science, 8*(3), 204-210.

Tompkins, C. A. (1994). *Right hemisphere communication disorders.* San Diego: Singular Publishing Group.

Tompkins, C. A., Baumgaertner, A., Lehman, M. T., & Fossett, T. R. D. (1997). Suppression and discourse comprehension in right brain-damaged adults: A preliminary report. *Aphasiology, 11*, 505-519.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Trivers, R. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology, 46*, 35-57.

Tulving, E., Kapur, S., Craik, F. I. M., Moscovitch, M., & Houle, S. (1994). Hemispheric encoding/retrieval asymmetry in episodic memory: Positron emission tomography findings. *Proceedings of the National Academy of Sciences, 9*, 2016-2020.

VanLancker, D. R., & Kempler, D. (1987). Comprehension of familiar phrases by left but not by right hemisphere damaged patients. *Brain and Language, 32*, 265-277.

Warburton, E., Wise, R. J., Price, C. J., Weiller, C., Hadar, U., Ramsay, S., & Frackowiak, R. S. (1996). Noun and verb retrieval by normal subjects: studies with PET. *Brain, 119*, 157-179.

Wellman, H., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition, 30*, 239-277.

Whiten, A. (1996). When does smart behaviour-reading become mind-reading? In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 277-292). Cambridge, UK: Cambridge University Press.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103-128.

Winner, E., Brownell, H., Happé, F., Blum, A., & Pincus, D. (1998). Distinguishing lies from jokes: Theory of mind deficits and discourse interpretation in right hemisphere brain-damaged patients. *Brain and Language, 62*, 89-106.

Winner, E., & Gardner, H. (1977). The comprehension of metaphor in brain damaged patients. *Brain, 100*, 717-729.

Zaitchik, D. (1991) Is only seeing really believing? Sources of the true belief in the false belief task. *Cognitive Development, 62*, 91-103.