The Social Brain
Evolution and Pathology

Editors Martin Brüne Hedda Ribbert Wulf Schiefenhövel

WILEY

The Social Brain

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

The Social Brain
Evolution and Pathology

Edited by

Martin Brüne
Centre for Psychiatry and Psychotherapy, University of Bochum, Germany

Hedda Ribbert
Centre for Psychiatry and Psychotherapy, University of Bochum, Germany

Wulf Schiefenhövel
Human Ethology Group, Max-Planck-Society, Andechs
Centre for Human Sciences, University of Munich, Germany

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Painting "social chess" by Avelke Edel (artist), Bochum, Germany (www.edel-art.de)

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0 470 84960 6

Typeset in 10/12pt Times by Dobbie Typesetting Ltd, Tavistock, Devon
Printed and bound in Great Britain by Biddles Ltd, Guildford
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

Contents

List of Contributors

JESSE M. BERING	*Department of Psychology, University of Arkansas, Fayetteville, AR 72701, USA*
DAVID F. BJORKLUND	*Department of Psychology, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, USA*
JESSICA BRAMHAM	*Neuropsychology Unit, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, UK*
LESLIE BROTHERS	*2045 Federal Avenue, Los Angeles, CA 90025, USA*
MARTIN BRÜNE	*Centre for Psychiatry and Psychotherapy, University of Bochum, Alexandrinenstrasse 1–3, D-44791 Bochum, Germany*
RICHARD W. BYRNE	*School of Psychology, University of St Andrews, St Andrews, Fife KY16 9JU, UK*
BRUCE G. CHARLTON	*Department of Psychology, University of Newcastle upon Tyne NE1 7RU, UK*
KATARZYNA CHAWARSKA	*Child Study Center, Yale University School of Medicine, PO Box 207900, New Haven, CT 06520, USA*
TIM J. CROW	*University Department of Psychiatry, Warneford Hospital, Oxford OX3 7JX, UK*
GERHARD DAMMANN	*Department of Psychiatry and Psychotherapy, University of Basel, Wilhelm Klein-Strasse 27, CH-4025 Basel, Switzerland*
MARIA LAURA GARAU	*Buenos Aires Neuropsychiatric Center, Ciudad de la Paz 1808, 4th Floor, 1428 Buenos Aires, Argentina*

WARREN JONES	*Child Study Center, Yale University School of Medicine, PO Box 207900, New Haven, CT 06520, USA*
INGRID E. JOSEPHS	*Department of Clinical Psychology and Personality, University of Nijmegen, 6500 HE Nijmegen, The Netherlands*
WINFRIED KAIN	*Department of Psychology, University of Salzburg, Hellbrunnerstrasse 34, A-5020 Salzburg, Austria*
PETER KINDERMAN	*Department of Clinical Psychology, University of Liverpool, Whelan Building, Quadrangle, Brownlow Hill, Liverpool L69 3GB, UK*
STUART KINNER	*School of Psychology, University of Queensland, Brisbane, Queensland 4072, Australia*
AMI KLIN	*Child Study Center, Yale University School of Medicine, PO Box 207900, New Haven, CT 06520, USA*
WILLIAM C. MCGREW	*Department of Anthropology and Zoology, Miami University, Oxford, OH 45056, USA*
LINDA MEALEY	*Psychology Department, College of St Benedict, St Joseph, MN 56374, USA*
ROBIN G. MORRIS	*Neuropsychology Unit, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, UK*
JOSEF PERNER	*Department of Psychology, University of Salzburg, Hellbrunnerstrasse 34, A-5020 Salzburg, Austria*
HEDDA RIBBERT	*Centre for Psychiatry and Psychotherapy, University of Bochum, Alexandrienstrasse 1–3, D-44791 Bochum, Germany*
GERHARD ROTH	*Brain Research Institute, University of Bremen, and Hanse Institute for Advanced Study, Lehmkuhlenbusch 4, D-27753 Delmenhorst, Germany*
ANDREA ROWE	*Neuropsychology Unit, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, UK*
TAMARA RUSSELL	*Section of Neuroscience and Emotion, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, UK*

WULF SCHIEFENHÖVEL	*Human Ethology Group, Max-Planck-Society, Von-der-Tann-Strasse 3, D-82346 Andechs, Germany*
ROBERT T. SCHULTZ	*Child Study Center, Yale University School of Medicine, PO Box 207900, New Haven, CT 06520, USA*
TONMOY SHARMA	*Clinical Neuroscience Research Centre, Stonehouse Hospital, 7 Twisleton Court, Priory Hill, Dartford, Kent DA2 6AU, UK*
SERGIO E. STARKSTEIN	*Buenos Aires Neuropsychiatric Center, Ciudad de la Paz 1808, 4th Floor, 1428 Buenos Aires, Argentina*
CARMEN STRUNGARU	*Department of Animal Physiology, Faculty of Biology, University of Bucharest, Splaiul Independentei 91-95, RO-76201 Bucharest, Romania*
FRED R. VOLKMAR	*Child Study Center, Yale University School of Medicine, PO Box 207900, New Haven, CT 06520, USA*

When editing the final draft of this book we received the very sad message that our colleague Professor Linda Mealey had passed away on November 5 at the age of 46. Until her untimely death, which she fought with the spirit so typical of her, Linda Mealey was a very active and highly respected member of the scientific community. The International Society for Human Ethology (ISHE) has created a new Linda Mealey Award for Junior Scientists in her honour.

Preface

The idea of compiling this volume has many fathers. Coming from different fields — psychiatry, psychology and human ethology — yet united in an evolutionary view on human perception, emotion, cognition and behaviour, we had planned and convened the international conference, 'The Social Brain — Evolution and Pathology'. Our aim was to open a forum facilitating communication and exchange between a number of disciplines concerned with the function and dysfunction of the very crucial human capacity of successfully interacting with conspecifics. The conference was held at the Centre of Psychiatry and Psychotherapy, University of Bochum, in late autumn 2000 and was dedicated to one of the pioneers of evolutionary psychiatry, Detlev Ploog, on the occasion of his 80th birthday in November 2000. Since the 1950s, probably influenced by his teacher Ernst Kretschmer, Detlev Ploog had conducted numerous experiments on the social behaviour of non-human primates at the Max-Planck-Institute for Psychiatry in Munich. He was clearly among the first psychiatrists who recognised the enormous impact of social interactions for the functioning of the human psyche. Moreover, he is probably the only German scholar who contributed chapters on ethology and evolutionary psychopathology to the multiple volume, *Contemporary Psychiatry* (originally published in German as *Psychiatrie der Gegenwart*) 35 years apart.

The label 'the social brain' characterises an essential part of our evolutionary history, because it is very likely that our being *animaux sociale* has shaped our emotional and cognitive brain mechanisms in very decisive ways.

Interestingly, researchers have neglected the significance of social aspects of cognition in non-human primates and humans for a long time. The traditional position was based, for example, on Wolfgang Köhler's famous experimental studies on tool use by great apes and assumed that the superior cognitive performance in humans was due to the demands of 'technological' processes, such as tool making and the solving of 'physical' problems. During the past decades, however, scientists such as Alison Jolly, Nicholas Humphrey, Richard

Byrne and Andrew Whiten, to name just a few, have recognised the impact of social life for the evolution of primate and human intelligence. Yet only recently have systematic studies been conducted to explore conditions that are characterised by a breakdown of social cognition.

In the texts of this multi-authored book, which offers new perspectives on social cognition, we set out by trying to reconstruct the evolution of social cognition in animals, especially our closest relatives, the great apes. We then discuss the evolution of culture, an exceptional element of human social life, and its reciprocal interaction with our brain. Finally, we focus on psychopathological conditions, which can, in many instances, be understood as a functional disruption of brain mechanisms normally safeguarding social cognitive performance. Accordingly, the book is organised in four major sections pertaining to these topics.

We hope to reach a broad audience interested in the *conditio humana* and to stimulate readers' curiosity and explorative behaviour towards our inter-disciplinary approach to neuroscience and new paradigms in evolutionary psychiatry in general.

We cordially thank our colleagues, who have so readily contributed their expertise to the conference and this volume. We are also indebted to Charlotte Brabants and Layla Paggetti of John Wiley & Sons Ltd. for their encouragement and support in completing and editing this volume.

Martin Brüne, Hedda Ribbert and Wulf Schiefenhövel
October 2002

Introduction

LESLIE BROTHERS
Independent Scholar, Los Angeles, CA, USA

Social cognitive neuroscience is emerging as a new field. Its central principle is that human beings use specific neural-cognitive mechanisms to process signals of the intentions and dispositions of others. To put flesh on the bones of this broad concept, however, many questions have to be answered. Can we define the components of social cognition accurately and localize them to specific brain circuits? How did these components evolve? Exactly how does the brain's inborn social potential interact with the environment during development? These questions transcend the boundaries of any single discipline. They call for thoughtful empirical research and careful, critical thinking from within such fields as developmental psychology, clinical neuroscience, comparative anatomy and primatology. The contributors to this volume, in taking up these important questions, rise to the interdisciplinary challenge.

One task is to specify what is meant by 'signals of the intentions and dispositions of others'. Eye gaze direction and certain facial muscle configurations were presumably important to our primate ancestors. An evolutionarily old neural system, present in the normal human brain at birth, may prepare the human infant to respond to such signals. Building upon responses to the sights and sounds of faces, such a system may act as scaffolding for the accumulation of subsequent social experiences. Elaborated in response to what the environment provides, it ultimately produces complex, finely-tuned responses to entities like belief, irony, *faux pas* and emotional communication — the spectrum of responses called 'theory of mind'. Studies of brain activation during the viewing of facial expressions, such as those reviewed by Morris, Bramham and Rowe (Chapter 11), are pieces of the larger puzzle. However, we don't have a clear picture of how inborn responses to

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

primitive signals develop into theory of mind, or of how the mature brain processes the array of everyday social events.

Learning is obviously important; several contributors emphasize the role of social learning. Based on comparative studies, Byrne (Chapter 3) proposes that relative neocortex size predicts the use of tactical deception in primate groups, but that insight is not required. Instead, there may have been selection pressure on the neocortex to store memories of identities and past actions of others, and match them appropriately to the current setting, thus allowing for social manipulation through simple but extensive social learning. Bjorklund and Bering (Chapter 7) additionally suggest that in our evolutionary history, a growing capacity for behavioural inhibition may have have complemented more specific social processes. They propose that, to understand the evolution of human social cognition, prolonged development, large brains with great capacity for learning, and a complex social environment must all be considered together. Certainly, for the human infant to progress from merely noticing eyes and faces, to responding to such biologically arbitrary signals as invitations to pretend play, jokes, and the gamut of interactive rituals, complex learning is key.

Another approach to dissecting the neural-cognitive mechanisms of sociality is the study of clinical entities. As Volkmar *et al.* (Chapter 9) point out, the spectrum of autistic disorders is heterogeneous and requires careful specification. However, autism clearly deserves its paradigmatic status in the field of social cognition. Autistic persons appear to have an inborn defect in elementary social processing, as shown by a number of studies of face and eye gaze processing. Schizophrenia, on the other hand, as Brüne (Chapter 13) demonstrates, is characterized by defects in decoding more subtle social signals. He argues convincingly that the pattern of social cognitive breakdown in schizophrenia seems to be the reverse of what is found in normal ontogeny. Both kinds of pathology affect theory of mind operations, but a breakdown in neocortical processes may be responsible for schizophrenic symptoms, whereas pathology in phylogenetically older structures may cause autistic deficits. While undoubtedly oversimplified, this scheme does justice to presumed ontogenetic interactions between basic social signal processors and neocortical mechanisms for efficient social learning, as well as the expected impact of individual genetic, intrauterine, and other environmental factors on both.

Other disorders of social cognition provide fertile ground for imaginative syntheses. Kinderman's chapter (Chapter 15) on paranoia and theory of mind pathologies, and Charlton's (Chapter 14) on the somatic basis of theory of mind, are two examples of the ways in which clinical observation and theoretical ideas can be mutually enriching. The social nature of the self, a concept explained by Josephs and Ribbert (Chapter 8), challenges our usual ways of thinking about that mysterious entity. Mealey and Kinner (Chapter 16) offer a compelling account of sociopathy. Dammann (Chapter 17) places

borderline personality disorder at the intersection of evolutionary psychology, attachment and mentalising.

While speculation is valuable, it is essential that emerging conclusions be challenged and corrected before they become entrenched. A strength of social cognitive neuroscience, reflected in this volume, is the fact that new theoretical ideas are repeatedly scrutinised in the light of empirical data. For example, Morris, Bramham, and Rowe (Chapter 11) show that two major categories of theory of mind — beliefs and emotions — don't produce anatomically separate patterns of activation in prefrontal cortex. Such careful analysis provides a good corrective to overly simple conceptions of theory of mind. Perner and Kain (Chapter 10) examine the relation between theory of mind deficits, frontal lobe pathology and executive function. They conclude that executive competence is not a prerequisite for theory of mind and offer an important caveat regarding the interpretation of functional imaging studies in general. Roth (Chapter 2), in a different vein, cautions us that the human brain is not unique.

Other controversies in the field include the roles of frontal vs. limbic regions in theory of mind; what counts as theory of mind in non-human primates and for that matter other species; and whether the neural substrates of face perception are in fact face-specific, to name a few. The answers to these and other 'social brain' questions are still being hammered out.

There are larger issues at stake as well. If we consider how extensively the human mind is embedded in collective social practices, a new approach to the mind–brain problem suggests itself. First, we proceed from brain to culture: dense, complex social signalling is the basis of human culture, probably mediated by representations with a collective dimension, as described by McGrew (Chapter 5). Next, we move from culture to mind: cultural categories are the vehicles through which mind is defined and enacted, as Schiefenhövel's accounts (Chapter 6) of cross-cultural communication illustrate so vividly. Social cognitive neuroscience allows us to insert the essential middle term, culture, into the mind–brain equation. This is in contrast to traditional, non-social cognitive neuroscience, which frequently attempts to reify aspects of mind directly in neural activity, without considering that our definitions of mind are in many cases social artifacts to begin with.

Of course, social brain studies have their own potential pitfalls. One lesson to be drawn from this volume, for example, is that our conceptualisations of theory of mind, attractive as they may be, are subject to doubt. Nevertheless, as researchers continue to dissect and study social cognition in human and non-human primates, they may yet capture the elusive prize of a satisfactory account of the relation between the brain and the mind. In the process, they are illuminating the aspect of us — our sociality — that most delights, perplexes and motivates us in our daily lives.

PART I

Evolutionary Aspects of the 'Social Brain'

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

1

Stereotypy vs. Plasticity in Vertebrate Cognition

CARMEN STRUNGARU

Department of Animal Physiology, Bucharest University, Romania

The classic concept of animal and human behaviour opposes stereotyped, 'fixed action patterns' in animals and a high level of behavioural plasticity in humans. According to Jensen (1980) there is an 'intelligence scale' which is superposed over the phylogenetic scale.

Despite a great amount of knowledge in the field of animal behaviour, there are still only few comparative studies on the cognitive abilities and corresponding behaviour patterns of vertebrates (Thorndike 1911; Harlow 1949; Hodos and Campbell 1969; Bitterman and Woodard 1975; Macphail 1982), many of them dealing with closely related species. This is mostly due to technical difficulties in finding a common base of comparison, as for instance an equal level of motivation (Mackintosh 1988) between largely unrelated species, but also due to a still existing gap between experimental and ethological (natural) approaches and goals. Yet, between not influencing at all and totally controlling the behaviour of animals in a laboratory setting, intermediate approaches can be utilised.

Comparing data on human behaviour in identical or very similar natural conditions with those derived from testing other vertebrates can also be beneficial for understanding general, possibly universal cognitive mechanisms. This contribution is focused on non-human experiments; results of a study testing human subjects were reported elsewhere (Strungaru 1995).

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

HYPOTHESIS

Vertebrate and human cognition are parts of a continuum, due to general, universal patterns of the central nervous system in processing information and making decisions. If this assumption is correct, then it could be argued that:

- Some common schemata must be at work in the way the environment is accessed and made familiar.
- Learning strategies should be framed by some general patterns that were the most successful in the process of natural selection.

In order to test this, studies on spontaneous and induced behaviour of several species of vertebrates were conducted.

SUBJECTS

- Amphibia: *Salamandra salamandra* (13).
- Reptilia: *Lacerta viridis* (12), *Emys orbicularis* (10).
- Aves: *Gallus domesticus* (10), *Anas* sp. (10), *Anser* sp. (10).
- Mammalia: *Mesocricetus mesocricetus* (14), *Rattus* sp. (39), *Felix catus* (8), *Cercopithecus aethiopicus* (2).

Not all species were used in all tests. In the learning set tests, only the subjects that learned the task were included in further testing. Some of the animal subjects (salamanders, lizards and aquatic turtles) were collected from their natural habitats and kept in laboratory conditions for at least 1 year prior to testing. Only those who showed good accommodation to captivity (reproductive activities, normal feeding, good general aspect and behaviour) were used. Birds were obtained from eggs artificially incubated in the laboratory. Rats and hamsters were part of the laboratory stock. Cats were born and grown in the laboratory. Monkeys were part of an experimental stock not yet used in medical research.

TESTING METHODOLOGY

Spontaneous behaviour:

Open field — for exploratory behaviour.

Detour by diving — an original test for incentive learning.

Induced behaviour:

Learning set technique, implying visual discrimination.

OPEN FIELD BEHAVIOUR

The classic test (Hall 1934) was used with adaptations of the testing devices to the subjects' dimensions. The floor surface of the open field (OF) was divided into 44 squares. Starting from general observations of the behaviour of humans and traditional laboratory animals, as well as domestic animals in unknown environments, I divided the OF surface into three areas: peripheral, semiperipheral and central, considering them as areas with different degrees of 'risk'. Previous studies on OF behaviour in rats under different schedules (Dobre *et al*. 1981; Constantinescu (Strungaru), Turcu and Dobre 1983; Strungaru 1988) discussed in detail new methods of differentiating between general locomotor activity and active cognitive exploration, as well as methods of measuring the level of emotional reactivity .

The OF was bordered by 50 cm high opaque walls and was illuminated homogeneously. Observation of the behaviour of the subjects was done through an inclined mirror, so that the subjects were not directly disturbed by the presence of the experimenter. Each subject was tested for 7 min/day for 3 consecutive days at the same time of the day for a given subject. The chickens and ducks, as highly social species, expressed a marked alert reaction to isolation at the beginning of the test, as well as intense vocalisation; the tests with them were therefore conducted over more days than in the case of the other species. The start point was located at a mid-point between two corners of the enclosure, directly at the wall.

The parameters recorded were:

- The number and type (peripheral, semiperipheral, central) of squares crossed; in the beginning the animals tended to move close to the wall and then gradually, with growing exploratory motivation, they ventured more into the semiperipheral and then into the central area (Figures 1.2–1.7).
- The number and quality of different general movements that led to a widening of the exploratory field: bipedalism (hamsters, rats, cats), head movements facilitating visual exploration (turtles, lizards, birds).
- Emotional reactivity (micturition, defecation, freezing, auto-grooming, vocalisation); given the variability of emotional expression in the species tested, the degree of emotional reactivity was evaluated only in terms of presence/absence and frequency of manifestation.

Data on exploratory behaviour were expressed as percentages of total movement. In this way it was possible to compare the exploratory behaviour of animal subjects with low velocity with those with high velocity, as I was interested in the quality of motion rather than its quantity: if subjects of two species moved across 16 and 60 squares, respectively, this probably represented a species-specific pattern (Figure 1.1), e.g. connected to their speed of

locomotion, but if in most cases across species a limited percentage (in no species exceeding one-third) of the movement took place in the central area of the OF (Figures 1.2–1.7), then one might speak of a particular, possibly universal, strategy of exploring, in this case venturing into the most 'risky' area.

In the case of aquatic turtles, the exploratory behaviour was observed in both terrestrial and aquatic OFs.

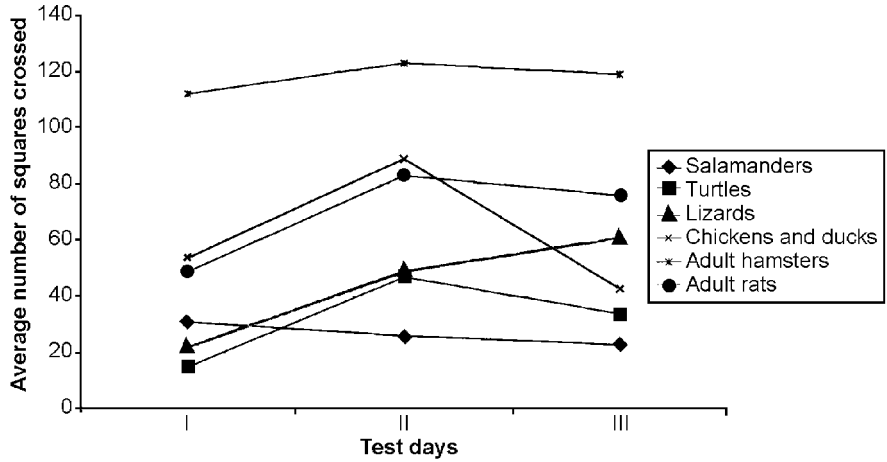


Figure 1.1. OF behaviour: average locomotor activity

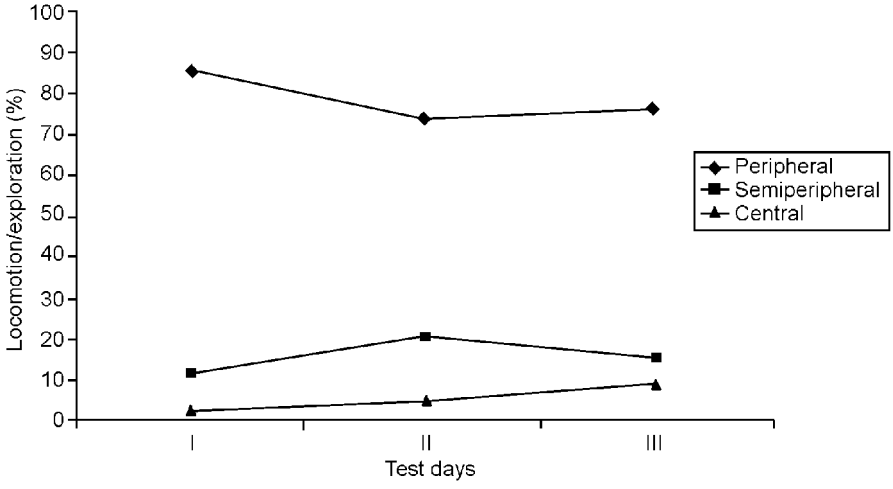


Figure 1.2. OF behaviour in salamanders

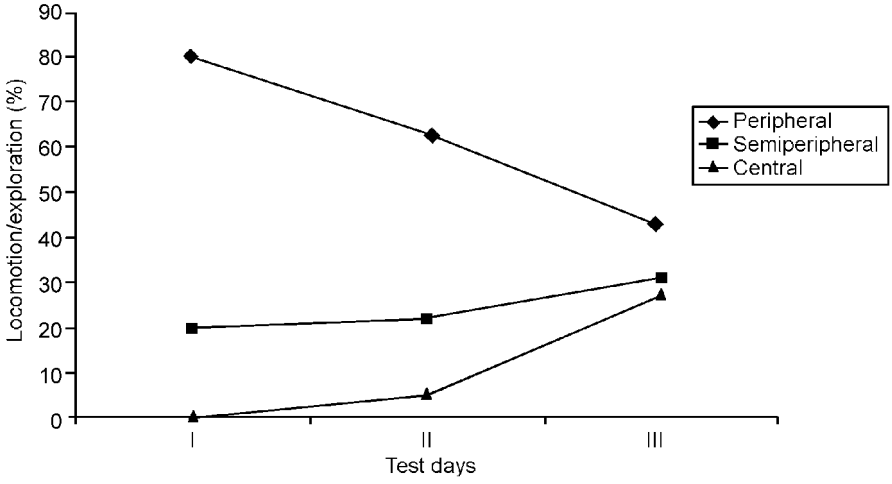


Figure 1.3. OF behaviour in lizards

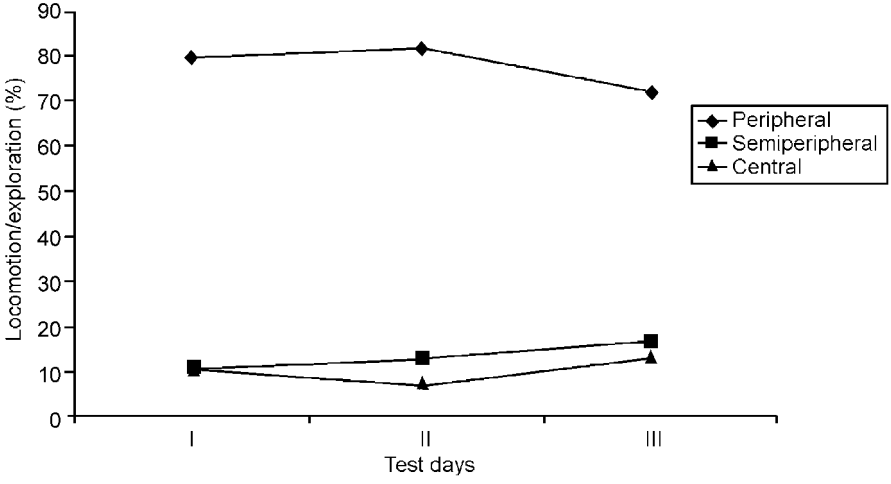


Figure 1.4. OF behaviour in turtles

Results

In the first day of OF testing, the subjects of the different species showed a low or very high level of locomotion (Figure 1.1). This activity had very little exploratory elements (central squares crossed, rearing, etc.) and was frequently accompanied by emotional signals (defecation, vocalisation, freezing, etc.). In many cases the animals froze in a corner of the OF. Only the hamsters ran

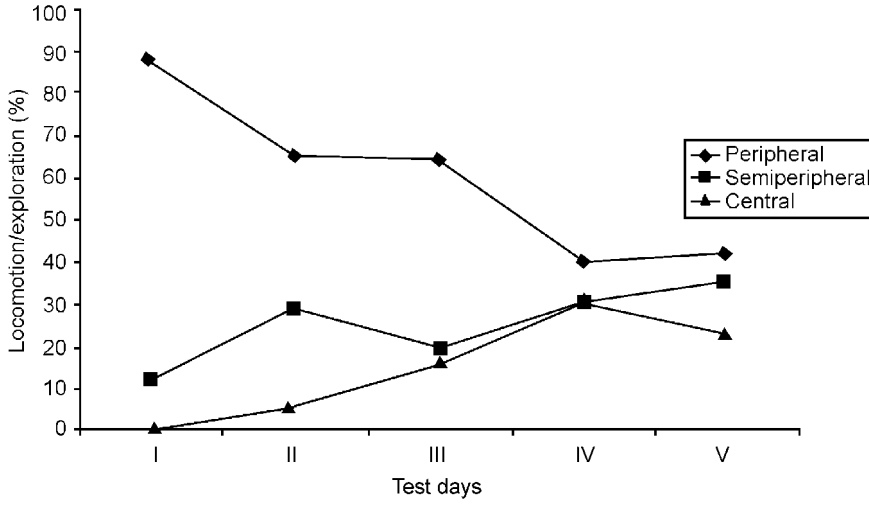


Figure 1.5. OF behaviour in birds

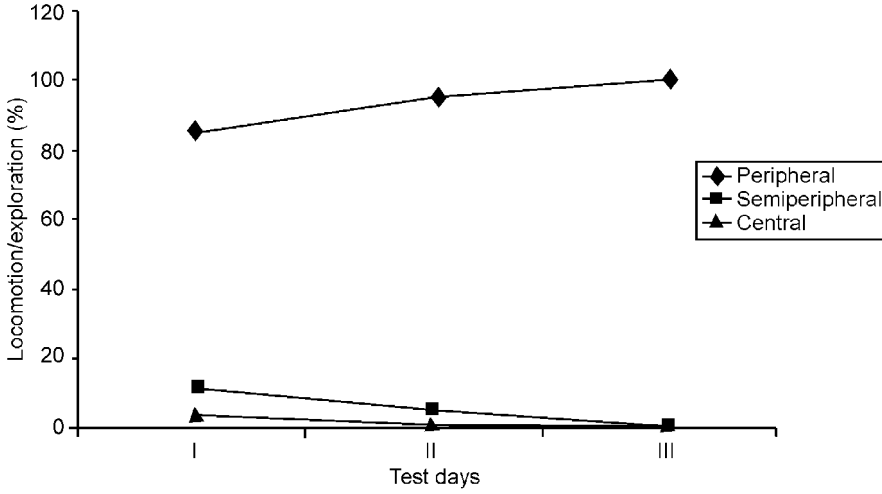


Figure 1.6. OF behaviour in hamsters

along the OF walls in search of an exit (Figure 1.6). Very few turtle individuals touched the central area of the terrestrial OF (Figure 1.4).

In the next 2 test days (with the exception of the hamsters, which showed ongoing escape behaviour and virtually no exploratory behaviour), the quality of the general behaviour was different, with less freezing, vocalising, micturition and defecation episodes and more visual and locomotor

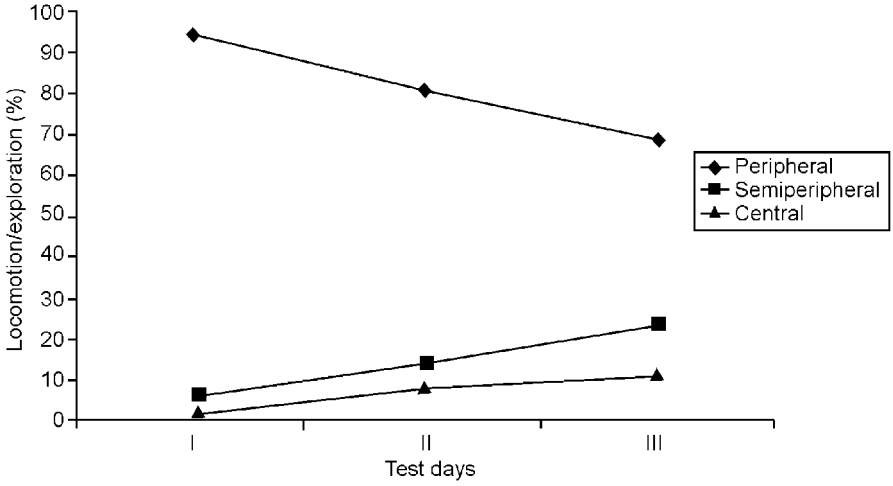


Figure 1.7. OF behaviour in rats

exploration. The peripheral area was explored first, with advances to and withdrawals from a corner apparently considered safe. In the process of realising that the new environment was not a dangerous one, the semi-central and central areas of the OF were also cautiously explored (Figures 1.2–1.7). In all the subjects tested except the hamsters, it was obvious that the same basic strategy was used.

SWIMMING TEST

In a water tank of $200 \times 80 \times 80$ cm, subjects learned to swim from a start platform to the opposite end, where an escape platform (rats) or a feeding platform (turtles, ducks, geese) was placed. All subjects rapidly learned to swim to the exit (food) platform. From one trial to next the speed increased significantly, while the stress signals decreased (Figures 1.8–1.10).

Interesting in this first stage of learning was the fact that each animal tested discovered and constantly used its own route and landmarks while swimming towards the target platform. Most of the rats avoided swimming in the middle of the water tank, preferring one of the lateral walls, and once they had reached the last landmark they changed direction towards the platform. They reached higher speed than the other species tested; this can be explained by a different motivation in their case (avoiding water) and by differences in species-specific environments.

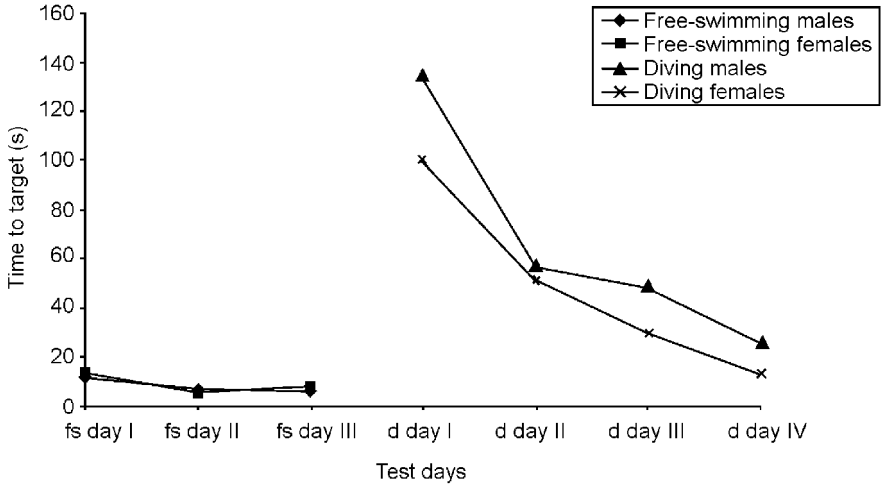


Figure 1.8. Detour problem in rats. fs, free-swimming; d, diving

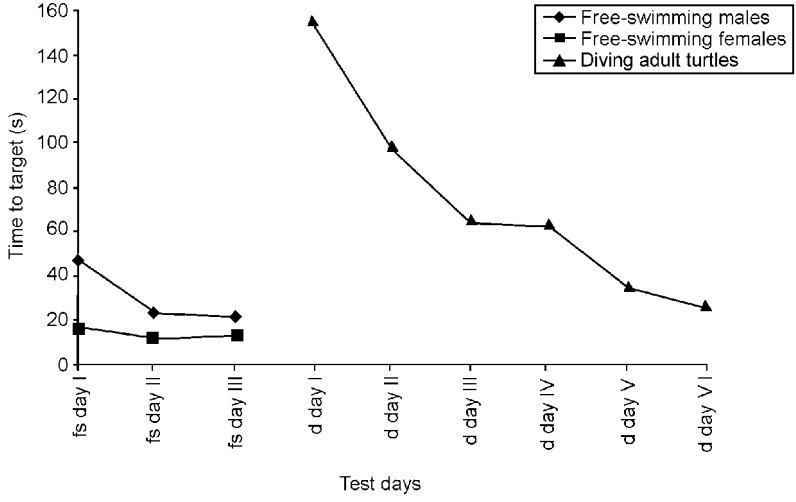


Figure 1.9. Detour problem in turtles. fs, free-swimming; d, diving

DETOUR BY DIVING TEST

After the learning criterion was reached, a transversally submerged transparent wall was fixed in the middle of the tank, barring free access to the target platform, which could now be reached only by diving down, at the subject's own initiative, to the bottom of the tank to pass under the wall. This new test

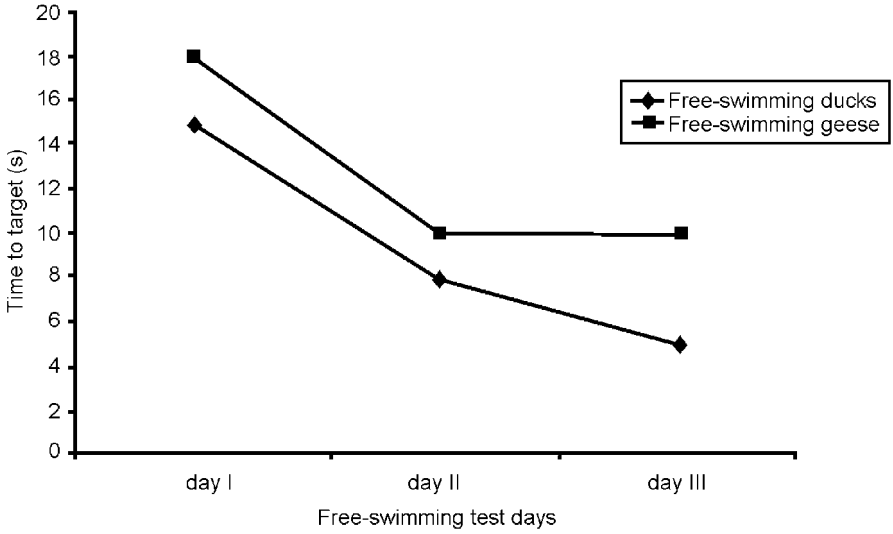


Figure 1.10. Detour problem in ducks and geese

was developed on the basis of observing rat behaviour in Porsolt, Bertin and Jalfre's (1977) 'despair behaviour' test, commonly used in experimental psychopharmacological studies, where the test is of great help for checking the antidepressive effects of drugs. In Porsolt's test, some rats, instead of giving up, were actively discovering the only possible way of escaping the 'swimming trap' by diving to the bottom of the small aquatic enclosure. Watching this, I was struck by the fact that only a few of the observed subjects did that, while the others continued to struggle at the surface of the water tank until they abandoned swimming.

Results

Rats

As terrestrial animals, their adaptation to an aquatic environment is mostly a matter of individual decision and level of ontogenetic development. Both adult and young rats (40 days old) were used. At the first trial with the submerged obstacle wall in the middle of the tank, all the rats showed the 'despair behaviour'. The adult individuals who dived and discovered the underwater exit kept this behaviour and, from the second trial onward, performed in an individual, highly stereotyped pattern (Figures 1.8, 1.11). It was obvious that they made use of spatial landmarks and of a 'moving-around' schema.

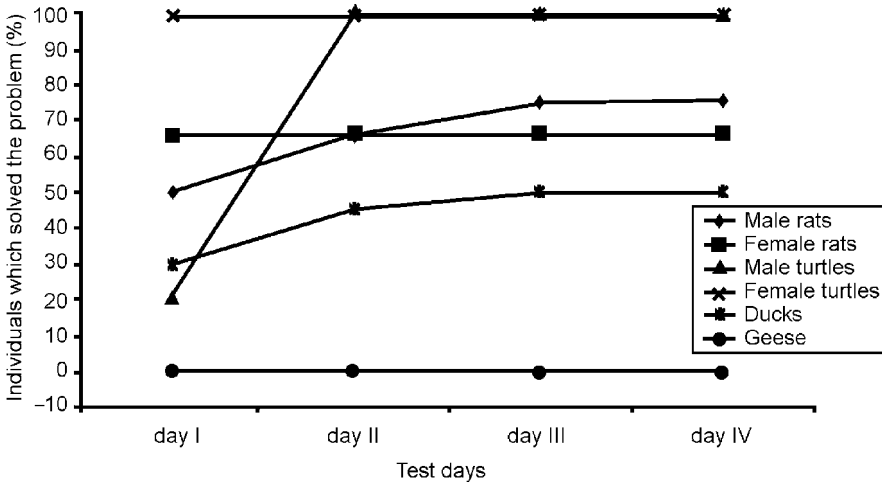


Figure 1.11. Detour problem performance

After an adult rat found its way to the safety platform by diving, for instance, along the left wall of the aquarium, it kept, in the next trials, this successful movement in the successful direction and did not search for other possibilities. None of the 15 young rats tested dived; all of them tried desperately to find an exit at the surface of the water. This result suggests age differences, due perhaps to the level of maturation of the CNS and/or individual life experiences.

The observed spontaneous 'going deeper into the water in order to escape the water' tells us about the individual cognitive and inventive abilities of animals. Adjusting previous behaviour to a new environmental situation depends on individual plasticity and pre-programmed abilities of the species.

Turtles

In aquatic turtle subjects, their 100% successful performance (Figure 1.11) reflects their high ability to adapt behaviourally in a familiar environment.

Ducks, geese

Tests with aquatic birds yielded prima facie surprising results, as most of them were not able to find the diving solution, even though diving for food is part of their natural behaviour. 50% of the ducks were able to discover the right answer to the test only when the submerged wall was lifted up to the water surface; however, the geese did not try to dive under the wall even in this

facilitating condition, therefore Figure 1.10 shows only the free swimming results for both species. For a possible explanation, see discussion below.

Finding and performing the right answer to the detour problem involved the same general pattern: appearance of an obstacle in an already learned task releases, in the first testing day, a confusion reaction evident in animals of all the species under study (Figures 1.8–1.10). The individuals able to find the way of overcoming the obstacle achieved this by trial and error in the first attempts, but later strictly by repeating the same psychomotor sequences and obviously being guided by the same exterior cues, showing an increased tendency to turn the successful behaviour into an stereotyped answer.

LEARNING SET TECHNIQUE

In tests selected for evaluating the capacity of learning and extracting rules from previous experiences, the positive alimentary reinforcement method was used. The experiment was conducted in a T-shaped maze and involved three stages:

- Spontaneous spatial learning of food location.
- Black vs. white discrimination.
- Triangle vs. square discrimination.

For each individual of each species, the number of trials was differentiated such that they could pass to the next test once they reached the learning criterion (2 consecutive days with maximum performance). In this way, overtraining was avoided.

Test 1: Simple Spatial Learning

In this test, animal subjects were free to make their directional choice. In order to avoid olfactory detection, the reward was delivered at either of the two arms of the maze, once the subject was already orientated to one direction. In the case of monkeys, which could not be handled outside their cages, the tests were done using cups situated at the left and right side of the cage door. Subjects could make their choice by touching one of the cups with a hand.

Each subject was exposed to a fixed number of trials, with no break between trials except the time to return it to the start point. The number of daily trials given to a subject was different from one species to another, in accordance with their specific locomotor speed and motivation level. The alimentary motivation was provided by partial food deprivation.

The parameters measured were:

- Start latency.
- Time to reach the target.

- Time to localise and start consuming the food.
- Direction selected.

Results

Turtles

They performed in a water T-maze. Given their high speed of movement in the aquatic environment, the starting point was on a small terrestrial platform. Thus, by measuring the start latency, the motivational level could be also evaluated. Each subject was exposed to six trials/day for a period of 5 consecutive days. As the group was heterogeneous in terms of age and sex, a separated analysis of the data was performed.

- *Start latency* diminished rapidly after the first testing day. There were no differences between age groups and only minor differences between the sexes, males being faster than females in this respect.
- *Time to reach the target* was shorter in young subjects than in the adults.
- *Time to localise the food*: especially in the first days, there were differences between the two time measurements, perhaps because it is more difficult for turtles to localise immobile food (Figure 1.12).
- *Direction selected*: despite the fact that each subject had the opportunity to find the food in both ends of the T-maze, 95% of the subjects selected the

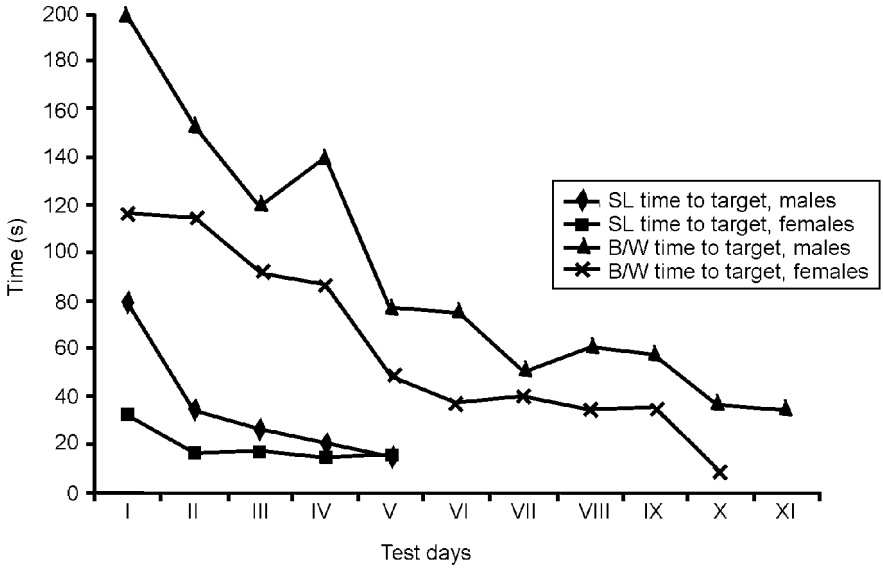


Figure 1.12. Time to target in simple learning (SL) and in black/white (B/W) discrimination in turtles

left end, even when they were placed on the right extremity of the start platform, thereby showing a strong tendency for directional perseverance.

Birds

As mentioned above, the chickens and ducks expressed a marked alert reaction to isolation at the beginning of the test, as well as intense vocalisation and lack of interest in food. Only from the third day on was it possible to expose each subject to 10 trials/day for 5 consecutive days. Once they had accommodated to the test situation, vocalisation and emotional defecation gradually diminished.

Behaviour differences between the two species were observed. In chickens, a constant directional perseveration was manifested during the entire testing period, while in ducks, directional perseveration was much less expressed. Ducks showed more intense exploration and their spontaneous directional alternation scores grew from one day to the next. It is noteworthy that the ducks did not alternate in a stereotyped way (e.g. left/right/left/right), but in a random manner.

Rats

The start latency was shorter in males than in females, with daily fluctuations in the latter, which might be related to the oestrus cycle and competition of motivations in females; time to reach target was not different in males and females, which may reflect a similar level of feeding motivation (Figure 1.13).

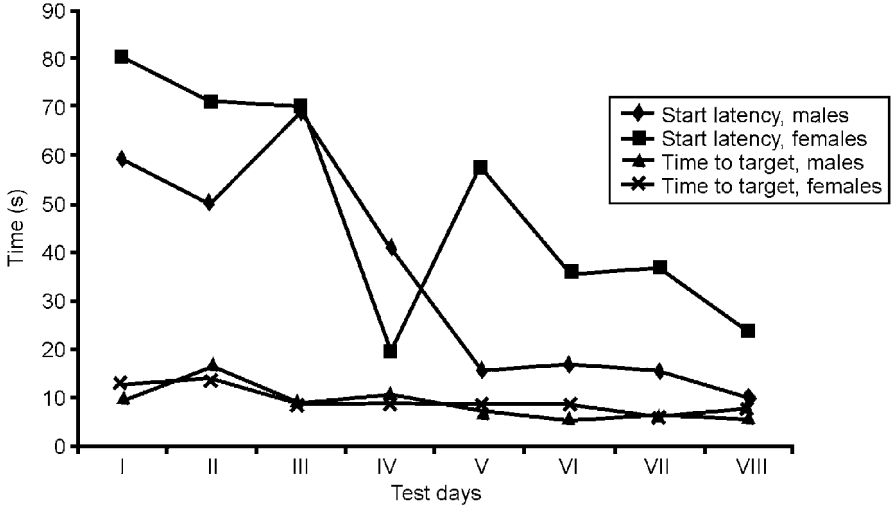


Figure 1.13. Simple learning in rats

Direction selected did not show perseveration or constant spontaneous alternation.

The subjects did not show negative emotional reactions to the test situation; on the contrary, many of them spontaneously returned to the start after consuming the food and waited for a new trial. This behaviour can be interpreted as a measure of the ability of rats to cope with very diverse environmental situations, which, among other traits, has made them the second-best adapted species in the world after humans.

Cats

Both adult and young cats were used. Because of their high speed of movement and reaction, the start latency could not be measured in this test and time to reach the target was no more than 2 s, despite the fact that the distance from the starting point was 5 m.

The parameter most interesting to analyse was direction preference. In kittens a direction perseveration of up to 96% was observed, while in some of the adult cats spontaneous alternation was observed. One possible explanation for this is that some orientational kinaesthetic and labyrinthian phenomena might be implied where the level of maturation is a key factor.

Monkeys

The number of subjects available for testing was too small for any statistics, but the observations collected can offer some information. In the free choice test, a slight directional perseveration can be described, without certainty that it is due to the animal's tendency or to its favourite position in the cage. As the monkeys were very sensitive to any type of stress, it was not possible to establish a clear score; often they simultaneously selected both targets, using both hands, and any attempt to correct them resulted in a refusal to return close to the door. It should be mentioned that these tests were conducted in the period 1991–1992, when clear legislation concerning animal protection was not yet in place in Romania.

Test 2: Black vs. White Discrimination

The test was carried out in the same ambience as the simple spatial learning test described above. In the new schedule, the two food targets were connected with visual signals, whose position changed randomly from one trial to another. For each species included in this series of tests, the position and dimension of the visual cues were designed such that they could be easily observed and utilised. The black target was the positive one for all the subjects under study.

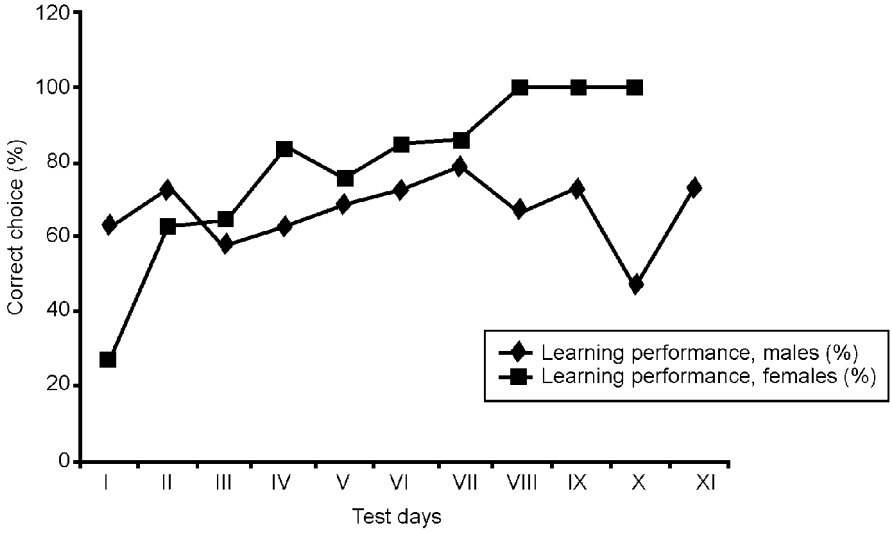


Figure 1.14. Black/white discrimination performance in turtles

Results

Turtles

Each subject was exposed to 10 trials/day for 10–14 consecutive days, i.e. until reaching the learning criterion. In the first days the performance was at chance level and the time to reach the target increased compared with the scores reached in the previous test (Figure 1.12). In my opinion, this reflects that the animals, detecting the environmental change, were more cautious and slightly disorientated by the fact that their presence in front of the feeding place was no longer rewarded every time. Starting with the third testing day (females) and the fifth day (males) this parameter decreased substantially. Females reached maximum performance (0 errors in 10 trials) on the eighth day, while male performance lagged behind (Figure 1.14).

Birds

After the first testing day, when the bird subjects were also more alert and confused by the change in ambience, the associative discrimination performance reached a level very close to maximum. The errors, that were recorded were due mainly to events in the testing room (e.g. noise produced by changing the position of the visual cues, or some mistakes in manipulating the animals). Ducks were faster than chickens in reaching the learning criterion (Figure 1.15).

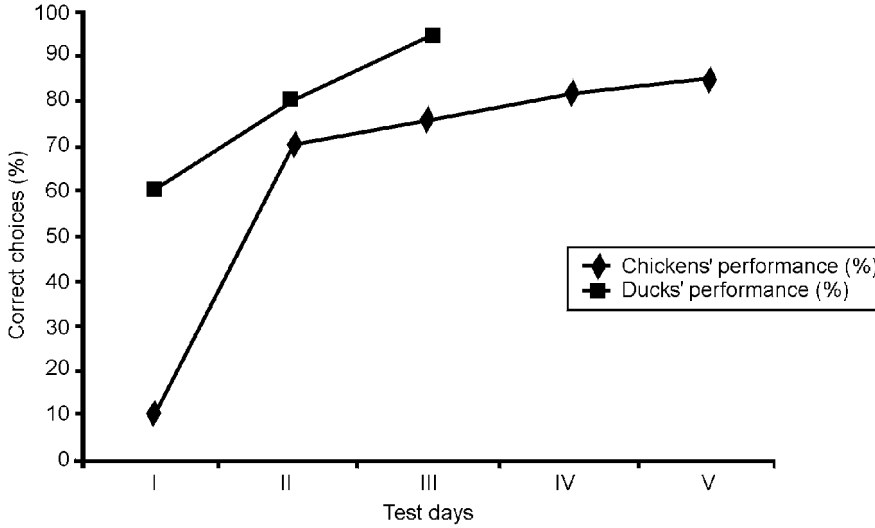


Figure 1.15. Black/white discrimination performance in birds

Cats

In both adult and young cats, individual differences could be described. These differences were related to their individual type of central nervous activity. Animals characterised by high velocity and a level of alertness made a greater number of errors and showed more directional perseverance than the slightly slower ones. The number of trials needed to reach the learning criterion was much higher than that needed by birds and monkeys.

Monkeys

As in the case of cats, individual behavioural differences were reflected in their performance. Subjects more active and less sensitive to the presence of the experimenter (subject D) made more mistakes. Once a subject realised that it had not indicated the correct target, it tried to reach the other target and obtain the reward; often it sanctioned the lack of reward with vocalisation and aggressive posture and facial expressions. The more 'shy' but still cooperative subject (SE), was more observant and made very precise, correct choices after the first few errors (Figure 1.16).

It is very important to take into account not only the pure learning performance of a subject but also the elements that facilitate or burden the learning process, if we are interested not only in how much animals can learn but also in how individuals do it. Usually, however, these kinds of differences do not lend themselves to statistical analysis (Warren 1968; Sutherland and Mackintosh 1971; Coppinger and Smith 1989).

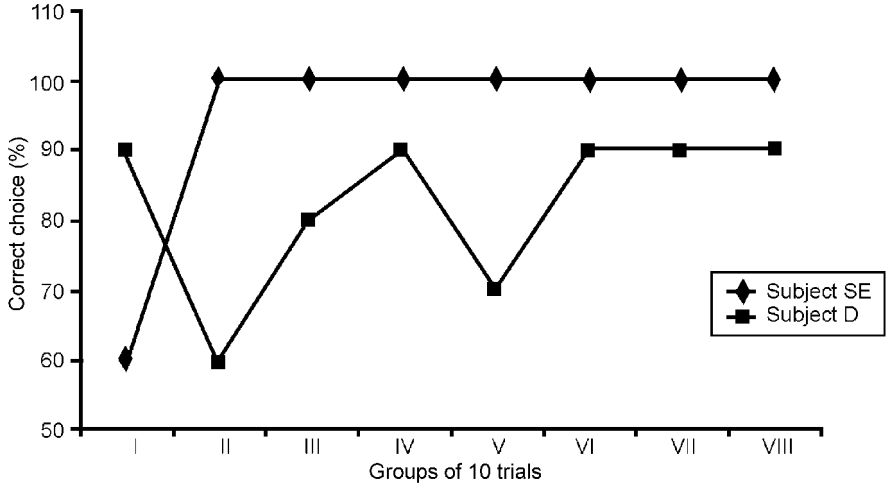


Figure 1.16. Individual performance in black/white discrimination in monkeys

Test 3: Discrimination of Geometric Shapes

Keeping the same general ambience, the black and white cues were replaced by cues with different geometric shapes: triangle and square. Both were black on a white background, and the triangle was the positive cue. As mentioned above, only subjects who reached the learning criterion in the black/white test were exposed to this new test.

Results

As in the previous test, changing the testing situation created confusion, expressed in a longer time to reach the target, emotional reactions, etc.

Turtles

At first they had difficulties in solving the new task, but from the 10th day of training onward 85% of the subjects were able to reach the learning criterion. There is no clear evidence that in the case of turtles we can talk about a real 'learning set' process but, nevertheless, the results show remarkable learning capacities in this vertebrate group (Figure 1.17). After a break period of 30 days I tested the subjects again in a 2 day long-term memory session. The results were impressive: many subjects reached the maximum performance from the first day on.

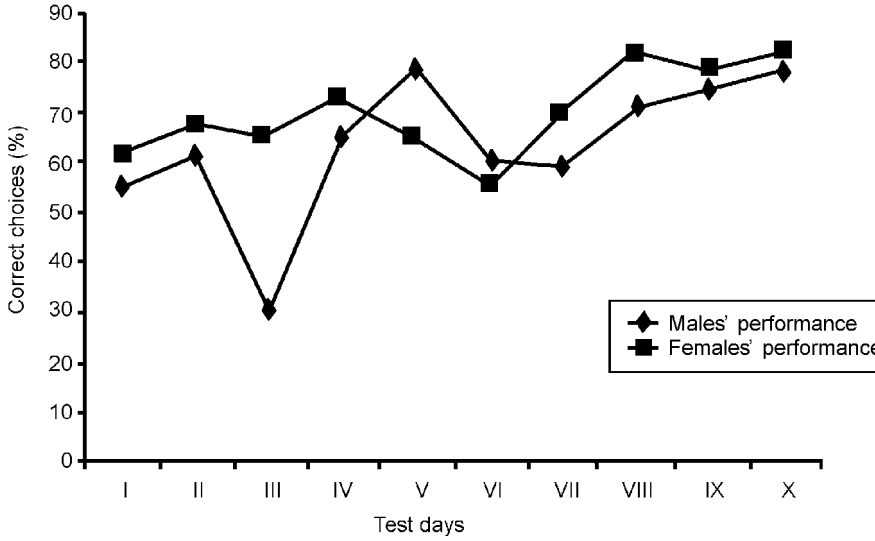


Figure 1.17. Geometric shapes discrimination in turtles

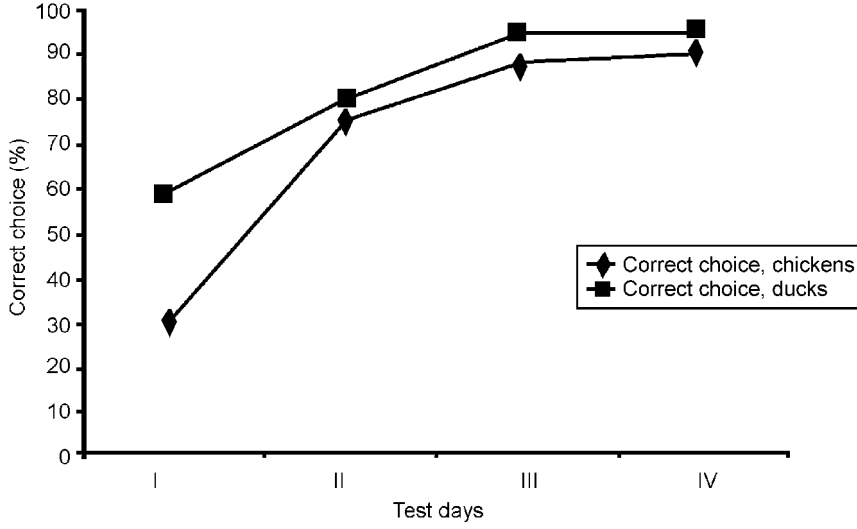


Figure 1.18. Geometric shapes discrimination in birds

Birds

Again, the performance of the bird subjects was surprising with regard to their speed and easy adjustment to the new test schedule, especially when compared with the performance of the mammals included in the experiment (Figure 1.18).

Cats

For the adult cats, the new test seemed to be very problematic. Most of them maintained their performance at chance level; one cat even developed 'experimental neurosis' and refused to search any longer for food after the first errors. Young cats showed more plasticity in adapting to the new experimental situation, but again, the data obtained do not support real 'learning to learn' evidence. What is important to underline in their behaviour is that once they realised that the learned system was no longer valid, most of them tried to solve the new problem using one of the stereotypes used previously — stereotyped alternation or directional perseverance.

Monkeys

Solved the triangle/square differentiation rapidly. Their performance was again depending on individual characteristics. Their behavioural answers depended on their capacity, or rather their incapacity, to control their precipitated motor reaction toward the food reward, even when they were fed *ad libitum*. One might say that proper inhibition was their problem, not motivation.

DISCUSSION

Exploration of the physical and social environment is the first ontogenetic step in cognitive development, at least in vertebrates. Natural environments can be portrayed as being in a *perpetuum mobile* state demanding behavioural plasticity. The animal mind can cope with this environmental plasticity by making use of successive stereotyped schemes that facilitate acquiring knowledge at lowest biological cost.

The data obtained in the study reported here indicate that these stereotypes are common in different classes of vertebrates, suggesting a universal cognitive pattern. There is always an equilibrium between the urge to know and the fear of the unknown (Montgomery and Monkman 1969; Cadland and Nagy 1969; Gray *et al.* 1981). This motivational conflict cannot be solved otherwise than step-by-step, advance and withdrawal, until the unknown becomes known. It is difficult to imagine, theoretically speaking, that a species whose members had adopted a different strategy could have had a phylogenetic future.

The 'intelligence scale' of Jensen (1980) is contradicted by the results reported in this contribution, which rather suggest a phylogenetic scale, with birds at the top. Intelligence or cognitive abilities are not monolithic processes. In humans it is well known that there are many types of performing in an intelligent way and of acquiring knowledge. Why should this not be the case also in the animal kingdom?

The high level of performance of the domestic birds used in my tests can be best explained by their species-specific perceptive and cognitive characteristics, which fit the visual discriminative association tasks. For any flying bird, it is always essential to detect the slightest visual details of their environment and to utilise them as landmarks in order to reach home. I am convinced that using other categories of cue stimuli would have dramatically changed the hierarchy of cognitive performance of the species tested.

In the same line of argument, the relatively bad performance of ducks and the completely lacking ability of geese to solve the detour-by-diving problem did not, in my opinion, reflect less cognitive capacity, but underlined the existence of other species-specific ways of solving these kinds of problems, such as flying or jumping over obstacles. For this reason I focused not on establishing any cognitive hierarchy but rather on finding out whether, despite the peculiarities of any given species in adapting to the environment, basic universal patterns could be detected. As Macphail (1982) underlines, there are no quantitative or qualitative differences between vertebrates in the process of learning; all of them show similar ways of habituation and solving conditioned learning tasks. The impression that some species are more intelligent than others might, in his view, be based on the different degrees of complexity and adaptability of perceptual and motor performance.

Animal behaviour, as well as human behaviour, is a successful mixture of plasticity and stereotypes. I should even dare to say that creating stereotypes is a necessary step in developing further behavioural plasticity.

REFERENCES

Bitterman, M.E. and Woodard, W.T. (1975). Vertebrate learning: common processes. In R.B. Masterson, M.E. Bitterman, C.E.G. Campbell and H. Hotton (eds), *The Evolution of the Brain and Behavior in Vertebrates*. Erlbaum: Hillsdale, NJ.

Cadland, D.K. and Nagy, M. (1969) The open field behaviour: some comparative data. In *Experimental Approaches to the Study of Emotional Behavior*. *Ann. N. Y. Acad. Sci.* **159**, 831–851.

Constantinescu (Strungaru), C.A., Turcu, E. and Dobre, V. (1983). Experimental psychopharmacological study on Aslavital for children (II): effect on open field behavior in Wistar rats. *Rom. J. Geront. Geriatrics* **4** (3), 207–217.

Coppinger, R.P. and Smith, C.R. (1989). A model for understanding the evolution of mammalian behavior. In H. Genoways (ed.), *Current Mammology*, vol. II. Plenum, New York, pp. 335–373.

Dobre, V., Constantinescu (Strungaru), C.A. and Turcu, E. (1981). The influence of age and Gerovital H3 treatment on the exploratory behavior and emotional reactivity in Wistar rats. *Rom. J. Geront. Geriat.* **2** (2), 313–320.

Gray, J.A., Davis, N., Feldon, J., Nicholas, J., Rawlins, P. and Owen, S.R. (1981). Animal models of anxiety. *Progr. Neuropsychopharmacol.* **5**, 143–157.

Hall, C.S. (1934). Emotional behavior in the rat. I. Defecation and urination as measures of individual differences in emotionality. *J. Comp. Psychol.* **18**, 385–403.

Harlow, H.F. (1949). The formation of learning sets. *Psychol. Rev.* **56**, 51–65.
Hodos, W. and Campbell, C.B.G. (1969). *Scala naturae*. Why there is no theory in comparative psychology. *Psychol. Rev.* **76**, 337–350.
Jensen, A.R. (1980). *Bias in Mental Testing*. Methuen, London.
Mackintosh, N.J. (1988). Approaches to the study of animal intelligence. *Br. J. Psychol.* **79**, 509–525.
Macphail, E. (1982). *Brain and Intelligence in Vertebrates*. Clarendon, Oxford.
Montgomery, K.C. and Monkman, J.A. (1969). The relation between fear and exploratory behavior. *J. Comp. Physiol. Psychol.* **68** (1), 107–112.
Porsolt, R.D., Bertin, A. and Jalfre, M. (1977). Behavioral despair in mice: screening test for antidepressants. *Arch. Int. Pharmacol. Therapeut.* **229**, 327–336.
Strungaru, C.A. (1988). The role played by some exogenous factors in behavioral differences of Wistar rats at different ages. *Abstr. Nat. Congr. Geront. Geriat. Bucharest*, **119**.
Strungaru, C.A. (1995). Comparative Ethological Mechanisms of Cognitive Processes. Doctoral Dissertation, Faculty of Biology, University of Bucharest.
Sutherland, N.S. and Mackintosh, N.J. (1971). *Mechanisms of Animal Discrimination Learning*. Academic Press, New York.
Thorndike, E.L. (1911). *Animal Intelligence*. Macmillan, New York.
Warren, J.M. (1968) Reversal learning and the function of learning sets by cats and rhesus monkey. *J. Comp. Physiol. Psychol.* **61**, 421–428.

2

Is the Human Brain Unique?

GERHARD ROTH

Brain Research Institute, University of Bremen, and Hanse Institute for Advanced Study, Delmenhorst, Germany

Humans are proud of their brain and their cognitive abilities, and many of us, including many neuroscientists, believe that the alleged uniqueness of human nature is due to the uniqueness of the human brain. Some popular claims about the human brain can be found, even in the scientific literature, e.g. (a) the human brain in general is anatomically unique; (b) humans have the largest brain in absolute terms; (c) humans have the largest brain relative to body size; (d) humans have the largest cerebral cortex, particularly prefrontal cortex; (e) humans have some brain centres or functions not found in other animals. These claims are briefly discussed below.

CLAIM (1) THE HUMAN BRAIN IN GENERAL IS ANATOMICALLY UNIQUE

This is completely wrong. All tetrapod vertebrates (amphibians, reptiles, birds, mammals) have brains that — despite enormous differences in outer appearance, overall size and relative size of major parts of the brain — are very similar in their general organization and even in many details (Wullimann 2000). More specifically, all tetrapod brains possess a median, medial and lateral reticular formation inside the medulla oblongata; a pons and ventral mesencephalon, including a noradrenergic locus coeruleus, serotonergic raphe nuclei and a medial ascending reticular activating system; there is a corpus striatum, a globus pallidus, a nucleus accumbens, a substantia nigra, a basal forebrain/septum and an amygdala within the ventral telencephalon; a lateral pallium,

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

homologous to the olfactory cortex of mammals; and a medial pallium, homologous to the hippocampal formation (at least Ammon's horn and subiculum). This means that all structures required for attention, declarative memory (or its equivalents in animals), emotions, motivation, guidance of voluntary actions and evaluation of actions are present in the tetrapod brain. These structures essentially have the same connectivity and distribution of transmitters, neuromodulators and neuropeptides in the different groups of tetrapods.

A more difficult problem is the presence of structures homologous to the mammalian isocortex in the telencephalon of other tetrapods. Amphibians possess a dorsal pallium, turtles and diapsid reptiles have a dorsal cortex plus a dorsal ventricular ridge (DVR), birds have a wulst and a DVR, and these structures are believed by many comparative neurobiologists to be homologous to the isocortex — and not to the basal ganglia — of mammals (Karten 1991; Northcutt and Kaas 1995; MacPhail 2000; Shimizu 2000). However, major differences exist between these structures with regard to cytoarchitecture and size. In amphibians, the dorsal pallium is small and unlaminated; in lizards it is relatively larger, and in turtles and some diapsid reptiles it shows a three-layered structure. In birds, those parts assumed to be homologous to the mammalian cortex (i.e. DVR and wulst) are large but unlaminated. In mammals, with the exception of insectivores and cetaceans, the dorsal pallium or isocortex shows the characteristic six-layered structure. Despite these differences it is safe to assume that the dorsal pallium and cortex of amphibians and reptiles is at least homologous to the limbic and associative cortex of mammals, while a primary sensory and motor cortex appears to be absent. When we compare birds such as pigeons or parrots with roughly equally intelligent mammals such as dogs, it then becomes apparent that the same or very similar cognitive functions are performed by anatomically very different kinds of pallium/cortex.

CLAIM (2) HUMANS HAVE THE LARGEST BRAIN IN ABSOLUTE TERMS

This is definitely wrong, as can be seen from Figure 2.1 and Table 2.1. Humans have large brains (1.3–1.4 kg average weight), which is the largest among extant primates (the extinct *Homo neanderthalensis* had a somewhat larger brain), but far from the largest among mammals. The largest mammalian brains (and of all animals) are found in elephants (up to 5.7 kg) and whales (up to 10 kg).

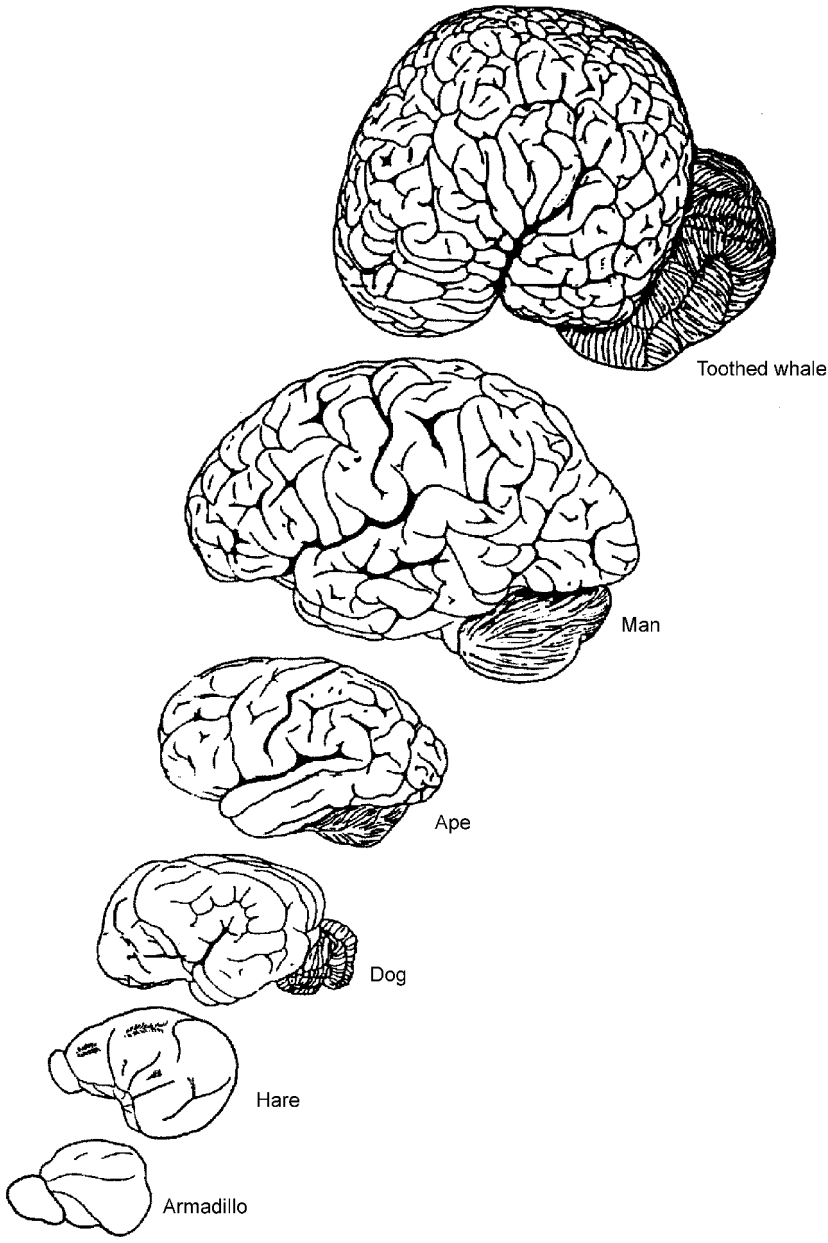


Figure 2.1. Series of mammalian brains, all drawn to the same scale. Evidently, man has neither the largest brain nor the most convoluted cortex. Convolution of the cortex as well as of the cerebellum increases monotonically with an increase in brain size

Table 2.1. Brain weight (g) in mammals

Sperm whale	8500
Elephant	5000
Man	1400
Horse	590
Gorilla	550
Cow	540
Chimpanzee	400
Lion	220
Dog	135
Cat	30
Rat	2
Mouse	0.4

CLAIM (3) HUMANS HAVE THE LARGEST BRAIN RELATIVE TO BODY SIZE

This is wrong, too. While the human brain occupies about 2% of body mass, in very small rodents relative brain size goes up to 10%. However, among primates, humans have the largest relative brain size. The relationship between brain size and body size has been discussed for more than 100 years (cf. Jerison 1973). It appears that body size is the single most important factor influencing brain size, i.e. large animals generally have large brains in absolute terms. However, increase in brain size does not strictly parallel the increase in body size, but follows only to the power of 0.66–0.75, i.e. two-thirds or three-quarters, depending on the statistics used (Jerison 1991, a phenomenon called *negative brain allometry* (Jerison 1973)) (Figures 2.2, 2.3). Consequently, small animals of a given taxon have *relatively* larger brains and large animals of this group have *relatively* smaller brains. Among mammals, this is reflected by the fact that in very small rodents the brain occupies up to 10% of body mass, in pigs 0.1% and in the blue whale, the largest living animal, 0.01% (Figure 2.4).

In addition, the different groups of vertebrates, while satisfying the principle of negative brain allometry, exhibit considerable differences in their fundamental brain–body relationship (Figure 2.5). Among tetrapods, mammals and birds generally have larger brains relative to body volume/weight than amphibians and reptiles, and among mammals, cetaceans and primates have relatively larger brains than other orders. Thus, during the evolution of birds and mammals, and more specifically of cetaceans and primates, genetic and epigenetic systems controlling brain size have undergone substantial changes in favour of relatively larger brains. These changes have resulted in enlargement of the brain beyond that associated with body size (Jerison 1991, 2000).

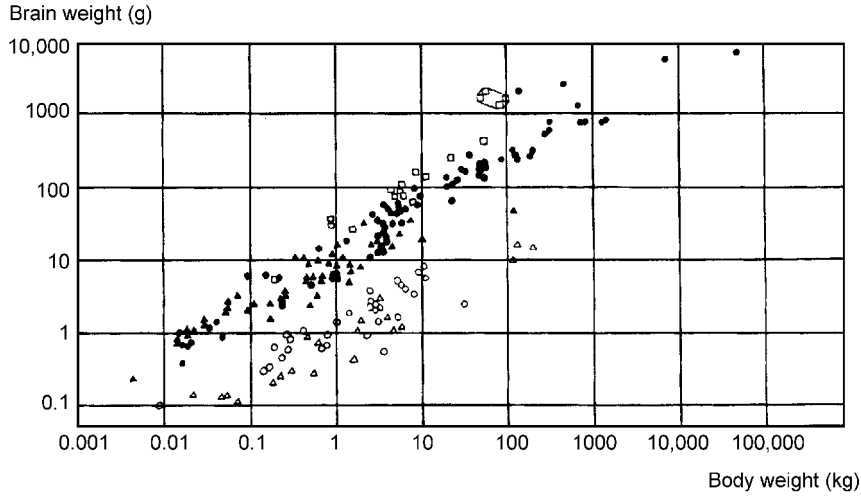


Figure 2.2. The relationship between brain size and body size in vertebrates (double-logarithmic graph). Open circles, bony fishes; open triangles, reptiles; filled triangles, birds; filled circles, mammals other than primates; open squares, primates; encircled open squares, *Homo sapiens*. After Jerison (1973)

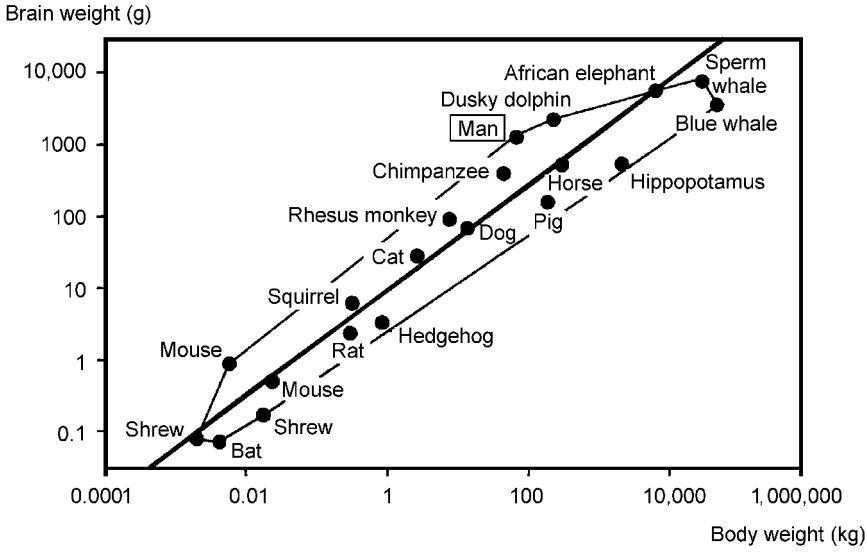


Figure 2.3. The relationship between brain size and body size in mammals. Data from 20 mammalian species (double-logarithmic graph). Modified from Nieuwenhuys, ten Donkelaar and Nicholson (1998)

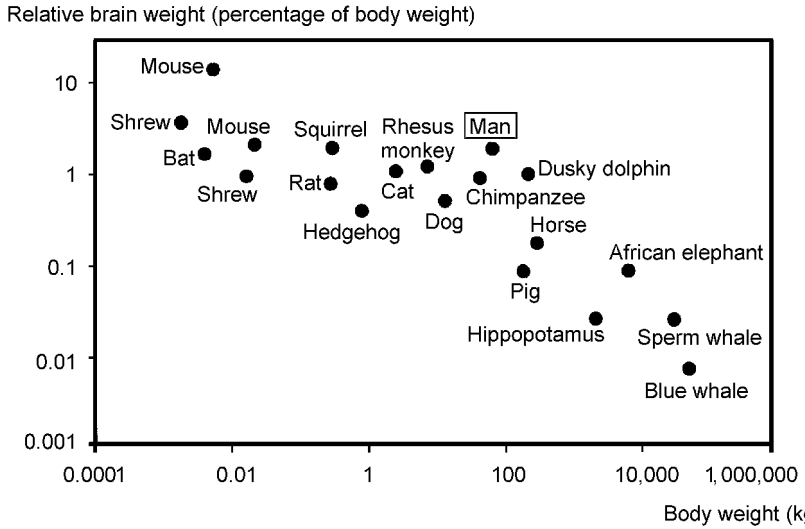


Figure 2.4. Brain weight as a percentage of body weight for the same 20 mammalian species as in Figure 2.3 (double-logarithmic graph). Modified from Nieuwenhuys, ten Donkelaar and Nicholson (1998)

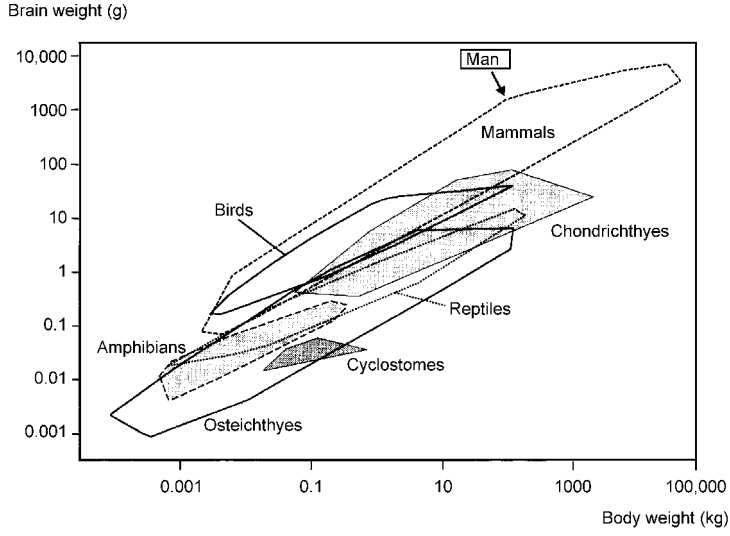


Figure 2.5. Diagrams showing the relationship between body weight and brain weight in the different classes of vertebrates (double logarithmic graph). Evidently, these classes differ in their general brain weight–body weight relationship, with the cyclostomes having the smallest and mammals having the largest relative brain weights. Remarkably, chondrichthyans (cartilaginous fishes, e.g. sharks and rays) have much larger relative brain weights than osteichthyans (bony fishes, above all teleosts). Modified after Jerison (1991)

Table 2.2. Encephalization quotients in mammals

Man	7.4	Marmot	1.7	Cat	1.0
Dolphin	5.3	Fox	1.6	Horse	0.9
Chimpanzee	2.5	Walrus	1.2	Sheep	0.8
Monkey	2.1	Camel	1.2	Mouse	0.5
Elephant	1.9	Dog	1.2	Rat	0.4
Whale	1.8	Squirrel	1.1	Rabbit	0.4

After Blinkov and Glezer (1968) and Jerison (1973, 1991).

Thus, contrary to a common belief, humans do not have the largest brain in either absolute or relative terms. Unless we accept that cetaceans and elephants are more intelligent than humans and/or have states of consciousness not present in humans, the absolute or relative size of the human brain *per se* cannot account for our actual or alleged superior cognitive abilities. However, among relatively large animals, man stands out with a brain that constitutes 2% of body mass. We can quantify this fact by determining the so-called encephalization quotient (EQ), which indicates the ratio between the actual relative brain size of a group of animals to the relative brain size as expected on the basis of brain allometry determined by body size alone (Table 2.2). Calculating the EQ for the human brain, it turns out that it is about seven times larger than that of an average mammal and about three times larger than that of a chimpanzee, if they were the same size as a human being (Jerison 1973, 1991).

While man stands out in this respect among primates, similar processes must have taken place among cetaceans. Toothed whales, particularly members of the family Delphinidae, exhibit EQs that are far superior to all primates except *Homo sapiens* (Marino 1998). While man has an EQ of about 7, the dolphins *Sotalia fluviatilis*, *Delphinus delphis* and *Tursiops truncatus* have EQs of 3.2, and the great apes (other than man) have EQs around 2. Thus, humans have a much larger brain than expected among primates, but even in this respect their brain is by no means unique, as the example of dolphins shows.

CLAIM (4) HUMANS HAVE THE LARGEST CEREBRAL CORTEX, PARTICULARLY PREFRONTAL CORTEX

There are enormous differences in both absolute and relative brain and pallial/cortical size among tetrapods and among mammals in particular. For example, man has a brain and a cortex that are roughly 3000 times larger in volume than those of a mouse. This implies that changes in *relative* size of cortex are inconspicuous, because in mammals cortical size rather strictly follows changes in brain size, but again, there are differences within mammalian groups. Apes

(including man) have somewhat larger isocortices than other primates and other mammals, because their forebrains (telencephalon plus diencephalon) are generally somewhat larger, constituting 74% of the entire brain as opposed to about 60% in other mammals, including mice. At 40% of brain mass, the human cortex has the size expected in a great ape (Jerison 1991).

The enormous increase in cortical volume is partly the result of an increase in brain volume, and consequently in cortical surface (which is related to an increase in brain volume by exactly the power of 2/3; Jerison 1973), and partly the result of an increase in the thickness of the cortex. The cortex is about 0.8 mm thick in mice and 2.5 mm in man. However, the number of neurons per unit cortical volume decreases with an increase in cortical thickness and brain size. While about 100,000 (or more) neurons are found in $1\,mm^3$ of motor cortex in mice, 'only' 10,000 neurons are found in the motor cortex of man (Jerison 1991). This decrease in the number of cortical neurons per unit volume is a consequence of a roughly equal increase in the length of axonal and dendritic appendages of neurons, in the number of glial cells and in the number of small blood vessels. Without such an increase in glial cells and blood vessels, large isocortices would probably be both architecturally and metabolically impossible.

Thus, the dramatic decrease in nerve cell packing density is at least partly compensated for by an increase in cortical thickness. This could explain why all mammals have a roughly equal number of neurons contained in a cortical column below a given surface area (e.g. $1\,mm^2$) (Rockel, Hiorns and Powell 1980). Furthermore, as explained above, what should count for the performance of neuronal networks is not so much the number of neurons *per se*, but the number of synapses their axons and dendrites form or carry, plus the degree of plasticity of synapses. An increase in length of axons and dendrites paralleling a decrease in nerve cell packing density should lead to more synapses, and such an increase in the number of synapses could also compensate for the strong decrease in nerve cell packing density. It has been estimated that the mouse cortex contains about 10 million (10^7) neurons and 80 billion (8×10^{10}) synapses and the human cortex about 100 billion (10^{11}) neurons and a quadrillion (10^{15}) synapses, 10,000 times more than the mouse cortex (Jerison 1991; Schüz and Palm 1989; Schüz 2000). These differences certainly have important consequences for differences in the performance of the respective cortices.

What about animals with brains and cortices that are much larger than those of man, e.g. elephants or most cetaceans? Shouldn't they be much more intelligent than man or have some superior states of consciousness (a popular assumption for whales and dolphins)? As to cetaceans, there is currently a debate on how many neurons their cortices really contain. Their cortex is unusually thin compared to large-sized land mammals and shows a different cytoarchitecture (e.g. lacking a distinct cortical layer IV). Accordingly, experts

report a lower number of nerve cells contained in a standard cortical column than in land mammals.

However, while Garey and Leuba (1986) report that in dolphins the number of cortical neurons per standard column is two-thirds that of land mammals, recently Güntürkün and von Fersen (1998), after examining the brains of three species of dolphins, reported that this value amounted to only one-quarter. Accepting this latter lower value, then — given a cortical surface of about 6000 cm^2 in dolphins (three times that of man) — the cortex of the bottlenose dolphin (*Tursiops truncatus*) should contain three-quarters the corresponding number of neurons found in humans, i.e. 6×10^{10}, which is about equal to the number of cortical neurons estimated for chimpanzees. Calculations of the number of cortical neurons in cetaceans with much larger brains and cortices, e.g. in the sperm whale with a cortical surface of more than 10,000 cm^2, are difficult, because precise data on cortical nerve cell number per standard cortical column are lacking. However, even assuming that, due to an enormous expansion of the cortex and consequent 'thinning out' of neurons, the respective value is only one-eighth of that found in land mammals, a sperm whale cortex should contain approximately the same number of cortical neurons as dolphins. Based on these calculations, we should expect cetaceans to be roughly as intelligent as non-human great apes, which is what cognitive behaviourists have discovered about these animals.

The case of elephants remains, with a similarly enormously large brain (around 4–5 kg) and a cortex of about 8000 cm^2 which, at the same time, is thicker than that of cetaceans but also possesses a typical six-layered structure. Assuming that the number of cortical neurons is two-thirds the value found in primates, elephants should have at least as many cortical neurons and cortical synapses as humans. Again, we do not know enough about the organization of the elephant cortex, but elephants should come close to the cognitive and mental capabilities of man, assuming that only the number of cortical neurons and synapses counted.

Perhaps it might be safer to restrict our consideration to the size of the associative cortex because, as mentioned at the outset, different kinds of consciousness are necessarily bound to the activity of specific parts of the associative cortex. There is a common belief that the associative cortex had increased dramatically in both absolute and relative terms during hominid brain evolution, and that this was the basis for the uniqueness of the human mind. However, such an increase is difficult to assess, as there are no precise criteria for distinguishing primary and secondary sensory cortical areas from true association areas. Recently, Kaas (1995) argued that the number of cortical areas increased dramatically from about 20 such areas in the hypothetical insectivore-like ancestor to more than 60 in primates. However, what has increased (according to Kaas) is the number of functionally intermediate areas (such as the visual area 3 and medial temporal lobe), but

neither the primary nor the highly associative areas. Kaas is right to warn about the danger of greatly underestimating the number of functionally different cortical areas in small-brained mammals.

Available data suggest that, contrary to common belief, the associative cortex has increased roughly in proportion to an increase in brain and cortical size. This is apparently the case for the prefrontal cortex (PFC), which is regarded by many neuroscientists and neurophilosophers as the true seat of consciousness. Anatomically, the PFC is defined as the cortical area with major (although not exclusive) input from the mediodorsal thalamic nucleus (Uylings and van Eden 1990; Roberts, Robbins and Weiskrantz 1998). Using this definition, it turns out that the PFC has increased isometrically with an increase in cortical and overall brain volume within groups of mammals, but here again we find an additional increase in relative PFC size, with an increase in absolute brain size across mammalian orders: in rats, the PFC constitutes 6.5%; in dogs, 8.7%; in cows 9.8%; and in man 10.6% of brain mass (Jerison 1997). What follows is that the human PFC has exactly the size expected according to primate brain allometry. Of course, cetaceans and elephants have prefrontal cortices that are much larger in absolute terms than the human PFC, but what they do with this massive 'highest' brain centre remains a mystery so far.

We have not yet found anything in brain anatomy that would explain the factual or alleged uniqueness of the human brain and of humans regarding cognition and consciousness. Given the fact that *Homo sapiens* has an absolutely and relatively large brain and cortex, it appears to be the animal with the highest number of cortical neurons and/or synapses, probably with the exception of the elephant. Thus, in this respect, humans are not truly exceptional. What is highly remarkable, however, is the strong increase in relative (and absolute) brain size in hominid evolution during the last 3–4 million years. While in the great apes and australopithecines, which were not our ancestors, brain size increases with body size to a power of 0.33, in the hominid lineage leading to *Homo sapiens* it increased to a power of 1.73, i.e. in a *positively allometric fashion*, which means that brain size increased faster than body size (Figure 2.6). However, the reasons for this phenomenon are completely unclear.

CLAIM (5): HUMANS HAVE SOME BRAIN CENTRES OR FUNCTIONS NOT FOUND IN OTHER ANIMALS

What remains is the question of whether there are any anatomical or physiological specialisations in the human cortex that could be correlated with the unique cognitive abilities attributed to man. As to the general cytoarchitecture of the human cortex, it is indistinguishable from that of other primates and most other mammals. Likewise, no differences have been

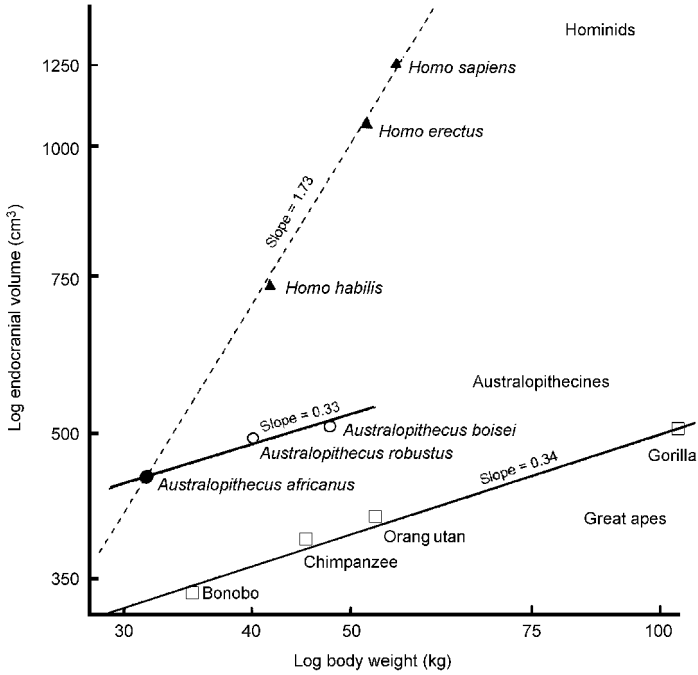


Figure 2.6. Increase in endocranial volume in the great apes, australopithecines and in hominids (double-logarithmic graph). Modified after Pilbeam and Gould (1974)

discovered so far between humans and non-human mammals with respect to the short- or long-term plasticity of cortical neurons, the action of neuromodulators, etc. Only two traits have been discovered that could drastically distinguish the human cortex from that of other primates: (a) differences in growth rate and length of growth period; (b) the presence of the Broca speech centre.

With regards to (a), maturation of the brain is more or less completed at 2 years after birth in prosimians and 6–7 years in monkeys and non-human apes, but the human brain still continues to mature until the age of 20, which is much longer than in any other primate (Pilbeam and Gould 1974; Hofman 2000). A critical phase in the development of the human brain seems to occur around the age of 2.5 years. At this time, major anatomical rearrangements in the associative cortex have come to a stop and the period of fine-wiring appears to start, particularly in layer III of the prefrontal cortex (Mrzljak *et al.* 1990). As mentioned above, at this time, human children 'take off' cognitively compared to non-human primates. Without any doubt, the drastically prolonged period of brain development constitutes one important basis for an increased capability of learning and memory formation.

Trait (b) concerns the presence of the Broca speech centre in the frontal lobe, responsible for temporal aspects of language, including syntax, along with the Wernicke speech centre in the temporal lobe, which is responsible for the meaning of words and sentences (although meaning is likewise dependent on syntax and grammar). It is to date unclear whether these speech centres are true evolutionary novelties. All mammals studied so far have a centre for intraspecific communication within the temporal lobe (mostly on the left side), which may be homologous to the Wernicke centre for semantics. It has been reported that destruction of these areas leads to deficits in intraspecific vocal communication (Heffner and Heffner 1995). In addition, it has long been argued that the posterior part (A 44) of the Broca speech centre in humans and the ventral premotor area of non-human primates are probably homologous (Preuss 1995). The ventral premotor area controls the movement of the forelimbs, face and mouth, which is likewise the case for the posterior portion of the Broca area.

According to a number of primatologists, non-human primates lack a direct connection between the motor cortex and the nucleus ambiguus, where the laryngeal motor neurons are situated. In man, bilateral destruction of the facial motor cortex abolishes the capacity to produce learned vocalization, including speech or humming a melody, while a similar destruction in monkeys has no such consequences (Jürgens 1995). According to a number of experts, the evolutionary basis for human language was an emotionally-driven stereotyped language typical of non-human primates. During hominid evolution, the cortex gained control over this system, such that beyond the initiation of hard-wired, innate sounds, a flexible production of sounds and their sequences became possible (Deacon 1990; Jürgens 1995). Such an interpretation, however, contrasts with recent evidence of a high degree of sound learning in monkeys (Zimmermann 1995) and the already mentioned consequences of destruction of left-hemispheric, Wernicke-like temporal areas in all mammals.

Be that as it may, non-human primates, including the great apes, are strongly limited, even in non-vocal speech based on the use of sign language or symbols, and these limitations seem mostly to concern syntax. Accordingly, anything concerning language in the human brain developed relatively recently or underwent substantial modifications; probably in the Broca centre rather than the Wernicke centre. Such an assumption is consistent with the fact that the most clear-cut differences between humans and non-human primates concern the syntactical complexity of language. Thus, during hominid evolution a reorganisation of the frontal-prefrontal cortex appears to have been organised such that the facial and oral motor cortices and the related subcortical speech centres came under the control of a kind of cortex that is specialised in all aspects of temporal sequence of events, including the sequence of action (Deacon 1990).

ACKNOWLEDGEMENT

I am grateful to Professor Jerison, UCLA/Hanse Institute for Advanced Study, for helpful criticism.

REFERENCES

Blinkov, S.M. and Glezer, I.I. (1968). *The Central Nervous System in Figures and Tables*. V.E.B. Fisher-Verlag, Jena.

Deacon, T.W. (1990). Rethinking mammalian brain evolution. *Am. Zoolog.* **30**, 629–705.

Garey, L.J. and Leuba, G. (1986). A quantitative study of neuronal and glial numerical density in the visual cortex of the bottlenosed dolphin: evidence for a specialized subarea and changes with age. *J. Comp. Neurol.* **247**, 491–496.

Güntürkün, O. and von Fersen, L. (1998). Of whales and myths: numerics of cetacean cortex. In N. Elsner and R. Wehner (eds), *New Neuroethology on the Move. Proceedings of the 26th Göttingen Neurobiology Conference*, vol. II. Thieme, Stuttgart. p. 493.

Heffner, H.E. and Heffner, R.S. (1995). Role of auditory cortex in the perception of vocalization by Japanese macaques. In E. Zimmermann, J.D. Newman and U. Jürgens (eds), *Current Topics in Primate Vocal Communication*. Plenum, New York, pp. 207–219.

Hofman, M.A. (2000). Evolution and complexity of the human brain: some organizing principles. In G. Roth and M.F. Wullimann (eds), *Brain Evolution and Cognition*. Wiley-Spektrum Akademischer Verlag, New York, pp. 501–521.

Jerison, H.J. (1973). *Evolution of the Brain and Intelligence*. Academic Press, New York.

Jerison, H.J. (1991). *Brain Size and the Evolution of Mind.* American Museum of Natural History, New York.

Jerison, H.J. (1997). Evolution of prefrontal cortex. In N.A. Krasnegor, G.R. Lyon and P.S. Goldman-Rakic (eds), *Development of the Prefrontal Cortex: Evolution, Neurobiology, and Behavior*. Brookes Publishing, Baltimore, MD, pp. 9–26.

Jerison, H.J. (2000). The evolution of neuronal and behavioral complexity. In G. Roth, and M.F. Wullimann (eds), *Brain Evolution and Cognition.* Wiley-Spektrum Akademischer Verlag, New York, pp. 523–553.

Jürgens, U. (1995). Neuronal control of vocal production in non-human and human primates. In E. Zimmermann, J.D. Newman and U. Jürgens (eds), *Current Topics in Primate Vocal Communication*. Plenum, New York, pp. 199–206.

Kaas, J.H. (1995). The evolution of isocortex. *Brain Behav. Evol.* **46**, 187–196.

Karten, H.J. (1991). Homology and evolutionary origins of the 'neocortex'. *Brain Behav. Evol.* **38**, 264–272.

Marino, L. (1998). A comparison of encephalization between odontocete cetaceans and anthropoid primates. *Brain Behav. Evol.* **51**, 230–238.

MacPhail, E.M. (2000). Conservation in the neurology and psychology of cognition. In G. Roth and M.F. Wullimann (eds), *Brain Evolution and Cognition*. Wiley-Spektrum Akademischer Verlag, New York, pp. 401–430.

Mrzljak, L., Uylings, H.B.M., van Eden, C.G. and Judás, M. (1990). Neuronal development in human prefrontal cortex in prenatal and postnatal stages. In H.B.M. Uylings, C.G. van Eden, J.P.C. de Bruin, M.A. Corner and M.G.P. Feenstra (eds),

The Prefrontal Cortex. Its Structure, Function and Pathology. Elsevier, Amsterdam, pp. 185–222.

Nieuwenhuys, R., ten Donkelaar, H.J. and Nicholson, C. (1998) *The Central Nervous System of Vertebrates*, vol. 3. Springer, Berlin.

Northcutt, R.G. and Kaas, J.H. (1995). The emergence and evolution of mammalian isocortex. *Trends Neurosci.* **18**, 373–379.

Pilbeam, D. and Gould, S.J. (1974). Size and scaling in human evolution. *Science* **186**, 892–901.

Preuss, T.M. (1995). Do rats have a prefrontal cortex? The Rose–Woolsey–Akert program reconsidered. *J. Cogn. Neurosci.* **7**, 1–24.

Roberts, A.C., Robbins, T.W. and Weiskrantz, L. (1998). *The Prefrontal Cortex. Executive and Cognitive Functions*. Oxford University Press, Oxford.

Rockel, A.J., Hiorns, W. and Powell, T.P.S. (1980). The basic uniformity in structure of the neocortex. *Brain* **103**, 221–244.

Shimizu, T. (2000). Evolution of the forebrain in tetrapods. In G. Roth and M.F. Wullimann (eds), *Brain Evolution and Cognition*. Wiley-Spektrum Akademischer Verlag, New York, pp. 135–184.

Schüz, A. (2000). What can the cerebral cortex do better than other parts of the brain? In G. Roth and M.F. Wullimann (eds), *Brain Evolution and Cognition*. Wiley-Spektrum Akademischer Verlag, New York, pp. 491–500.

Schüz, A. and Palm, G. (1989). Density of neurons and synapses in the cerebral cortex of the mouse. *J. Comp. Neurol.* **286**, 442–455.

Uylings, H.B.M. and van Eden, C.G. (1990). Qualitative and quantitative comparison of the prefrontal cortex in rat and in primates, including humans. In H.B.M. Uylings, C.G. van Eden, J.P.C. de Bruin, M.A. Corner and M.G.P. Feenstra (eds), *The Prefrontal Cortex. Its Structure, Function and Pathology*. Elsevier, Amsterdam, pp. 31–62.

Wullimann, M.F. (2000). Brain phenotypes and early regulatory genes: the Bauplan of the metazoan central nervous system. In G. Roth and M.F. Wullimann (eds), *Brain Evolution and Cognition*. Wiley-Spektrum Akademischer Verlag, New York, pp. 11–40.

Zimmermann, E. (1995). Loud calls in nocturnal prosimians: structure, evolution and ontogeny. In E. Zimmermann, J.D. Newman and U. Jürgens (eds), *Current Topics in Primate Vocal Communication*. Plenum, New York, pp. 47–72.

3

Tracing the Evolutionary Path of Cognition

RICHARD W. BYRNE
School of Psychology, University of St Andrews, UK

How can we discover the path along which modern human cognition evolved? Despite the thrill of uncovering human archaeology and the fascination of the detective work needed to interpret this evidence to best advantage, the preserved record gives precious few signs of hominid cognitive abilities. Fortunately, there is another way to find out about ancient minds — at least, there is for those human ancestors that we share with living non-human primates. That method can best be called *evolutionary reconstruction*. This chapter will begin by briefly sketching the principles of the method, then use some existing data on primate deception to illustrate its practice.

Let me begin with an analogy. Imagine a doctor studying haemophilia, a disease known to be carried by a single gene. An isolated population has come to light, in which a surprisingly large minority of individuals have haemophilia, and the question is when and how this state of affairs originated. First, the doctor will need to work out the family tree of the living members of the population, far enough into the past to find the point where the descent lines converge into one or a very few roots. If this can first be done, then tracing the origins of the haemophilia gene becomes possible, even if no historical record of its occurrence exists. Those who suffer from or carry haemophilia in the extant population are plotted as 'positives' on the terminal branches of the family tree. The inception of the disease can be identified as the first branch point from which all living members of this population *with* the affected gene are descended. This points to that single one of their ancestors who first suffered the genetic mutation, or carried the gene into the population from

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

elsewhere. Knowing 'who' and 'when' may therefore also give some clues as to 'how' it happened. In this not entirely fanciful example, the doctor uses the distribution of the characteristic in living people to reconstruct its origin in the past.

A comparable process can be applied to the early phases of human behavioural evolution. Instead of using individual people as the source of data, the *living species* of our animal relatives are studied. The importance of non-human primate behaviour for evolutionary psychology is the window it offers into human psychological traits, by using evolutionary reconstruction (of course, non-primate species are also related to humans, and their study can inform about even earlier periods of human ancestry). Before this was understood, living primates were studied for a different reason: in the simplistic and largely discredited search for species to 'model' particular human ancestors. In reality, it is very improbable that any living non-human primate closely resembles a human ancestor, whereas it is certain that we share inherited traits with our closest animal relatives by common descent from shared ancestors.

The equivalent of a family tree in evolutionary reconstruction is a phylogenetic classification of living species: a cladogram is constructed by grouping species that share clusters of derived characteristics (see Figure 3.1). Most reliable for this process are molecular characteristics, because the detail of complex biological molecules is highly unlikely to have become identical by chance or by convergent evolution (the risk of convergence is also frequently minimized by using parts of the DNA which do not appear to be expressed in the phenotype, such as pseudo-genes). For every junction on the cladogram, a real species must have existed, ancestral to all those living species that are grouped together at that level of similarity. That extinct species is called the 'crown group', and represents the most *recent* ancestor common to all the living forms. There will usually have been earlier species in the line of descent, also ancestral to those living species and to no others: they make up the 'stem group'. Cladistic taxonomy is a means of discovering information about crown group species, whereas many fossils may represent members of stem groups, or they may be of species on side branches that left no living descendants. For that reason, attributing a set of bones to a human ancestor is a risky business, whereas deducing the existence of extinct, crown group species is reliable, regardless of whether any of their fossils are ever found. Within limits, we can also obtain the dates at which these deduced ancestors lived, by calibration with fossils of known date. The calibration process depends on assuming that basic physical phenomena (e.g. concentrations of atmospheric gases, and levels of bombardment with cosmic rays) have remained essentially constant, which is quite likely. But it also assumes that a particular, dated set of bones can be unambiguously attributed to a crown group species on the cladogram, which is tricky. The dates of ancestor species are thus liable to occasional

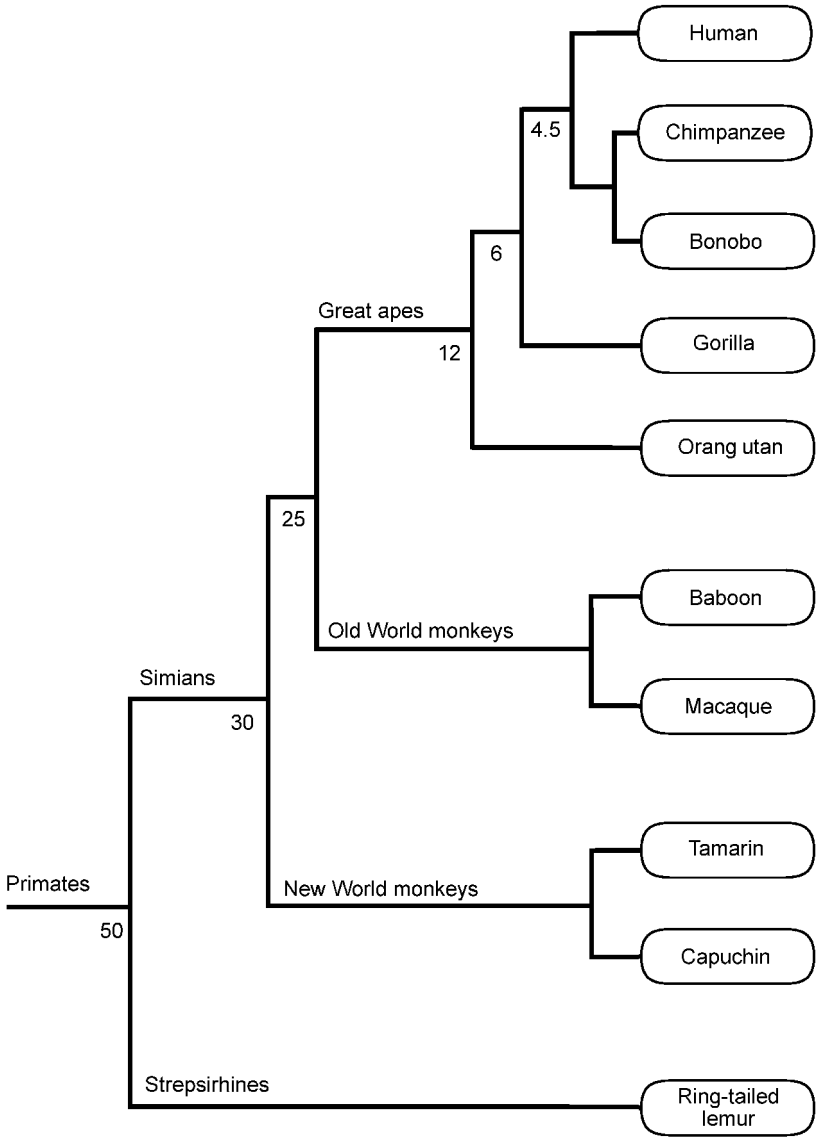


Figure 3.1. Phylogeny of the living primates. Only a few of the ca. 250 species are shown. Numbers represent the approximate estimates for the (crown group) ancestor population, for each clade that includes humans, in millions of years before the present, on the basis of calibration with the orangutan/human convergence at 12 million years ago.

reinterpretation, whereas the basic pattern of relationships among living species (and hence the inferred crown group species), when assembled from a wealth of evidence of different molecular characters, can be based firmly on data. In a well-studied group like the primates, the phylogenetic pattern is therefore unlikely to change.

Having deduced the existence of a series of ancestor species for (in this case) humans, we can go on to reconstruct the behaviour and mentality of these extinct species. For this process, the 'living evidence' from their modern descendants, the non-human primates and ourselves, is essential. To make reliable generalisations about the shared characteristics of a group of related species, it is important to have evidence from a range of species in the group: a single datum might be just a rogue exception. Unfortunately, traditional comparative psychology often used few species, often only three: the chimpanzee *Pan troglodytes*, which had to represent the four to six species of non-human great apes now recognized; the rhesus monkey *Macaca mulatta*, representing the 180 or so species of other primates; and the rat *Rattus norwegicus*, to represent non-primate mammals (over 4300 species; Corbet and Hill 1991). It was hard to escape the suspicion that a vestige of the *Scala Naturae* lurked behind comparisons of that sort! Today, with the wealth of behavioural evidence on non-human primates and many other animals, we should be able to do better — but evidence about animal mentality is still in rather short supply, because most of it still comes from captive studies. Laboratories, for inevitable reasons of practicality and cost, tend to have a limited range of species, most often the chimpanzee, the rhesus macaque or the capuchin monkey, *Cebus apella*, and, of course, the white, laboratory strain of the Norway rat. The problem is obvious. For some purposes, captive studies remain uniquely valuable: in particular, for testing predictions from theory by creating situations to order, that are rare or impossible in the field, with conveniently generated control data. However, to transcend the limitations that fatally beset much of traditional comparative psychology, we must go beyond the confines of the laboratory and accept evidence from a less restricted range of species in the field — even though that evidence may be harder to evaluate and control. In the example I will pursue in this chapter, evidence of primate deception under natural conditions comes from a wide range of taxa, including all the major groupings of the primate order (Byrne and Whiten 1992). For this reason, it is convenient to use it to illustrate the potential of the method of evolutionary reconstruction.

Deception in nature is widespread, and often has nothing to say about psychology. Consider those wonderful devices of animal camouflage, such as the plumage of the potoo *Nyctibius griseus*, a neotropical bird that resembles the surface of a dead branch so closely that it can spend the day perched in full view on a fence post, and yet remain unnoticed by a keen birdwatcher who inspects the fence. The potoo's deception is *strategic*: the bird has no other

option open to it, and no decision-making is involved beyond picking the site where it perches. Although potoos do usually perch on posts or branches that render them safe, when they occasionally choose inappropriately, they continue to behave as if they were invisible, and can then sometimes be touched by hand. In contrast to this strategic deception, consider the case of a female mountain gorilla, *Gorilla b. beringei*, living in a group with a dominating leader male. She might in fact sometimes prefer to mate with one of the other males in the group, perhaps a younger male who has some potential she can appreciate; but such copulation is not permitted by the leader. Careful observations of the mountain gorillas of Karisoke, Rwanda, show that some females can employ several tactics to achieve forbidden matings, and these tactics only succeed if they deceive the leader male (for full details, see records in Byrne and Whiten 1990). Pandora, a female in the silverback Beetsme's group, showed just such an evident preference for a younger male, Titus. Most simply, Pandora sometimes hung back when the group moved, so that she remained in Titus's company, and then mated out of sight of Beetsme. This might sometimes have happened by coincidence, but it was noted surprisingly often. More tellingly, Pandora sometimes actively solicited Titus, with side-to-side head flagging, and both went off alone for a few minutes and mated. If silverback leaders discover these 'secret' copulations, they attack and beat the female concerned: detection is punished. Unfortunately, gorillas, both male and female, normally emit copulation calls (probably homologous to human orgasm); this vocalization can easily betray the act of mating. It is of interest, then, that some but not all gorillas are able to refine their deception one step further: they can inhibit the normal copulation calls, mating silently. Both Pandora and Titus were typically silent when they mated. However, Shinda, a young male in another group, was not, although his contorted facial expressions when mating suggested that he was trying hard to do so. These observations raise several issues: did the perpetrators intend to create or preserve a false belief?; was the victim, the silverback, aware of the possibility that he might be a cuckold? At present, there is insufficient data to answer these interesting questions. However, the point to note is that the behaviour functioned if and only if another individual was deceived (the leader male). Moreover, the actions were certainly deployed as tactics: the gorillas took advantage of particular circumstances and behaved in ways that would have been unusual otherwise (e.g. mating in silence). The observations thus show *tactical deception*.

The full set of records of primate tactical deception, as published in 1990, comprises 253 records of behaviour (Byrne and Whiten 1990). Each one was recorded by an observer who was very familiar with the behaviour of the species concerned, a scientist working with a group of primates that had been habituated to close observation and whose members were known individually. Each observation had been unexpected when it occurred, and as such had been

carefully recorded, usually in longhand, in case some scientific use might one day be made of it. In a sense, then, these reports are anecdotal — they are narratives (*Concise Oxford English Dictionary*, 5th edn, 1964: Anecdote, n. Narrative of detached incident). However, 'anecdotal' in psychology usually has more negative connotations. It typically refers to badly-observed, second-hand stories, retailed by people who did not understand what they were seeing. In that highly pejorative sense, the published records of primate tactical deception are not anecdotes. The history of their collection and publication is revealing.

In 1983, my colleague Andrew Whiten and I had originally noticed several cases of tactical deception by the baboons whose socioecology we were studying. At that time, only chimpanzees had been recorded as using deception for social manipulation (Goodall 1971), so we felt these observations of sufficient interest to merit publication (Byrne and Whiten 1985), as they were apparently unique among monkeys. However, we soon discovered informally that other primatologists had in fact seen similar things happening, but had been shy of publishing 'anecdotes', fearing ridicule. Suspecting that an important phenomenon might have been overlooked wholesale, we contacted those primatologists whom we knew to be in a position to observe any tactical deception that might occur. These were primatologists whose observations were close-range and detailed and whose subjects were well habituated and known individually. We asked them to contribute any records that they considered to match our working definition of tactical deception. The resulting corpus was evidently incomplete, biased by our limited knowledge of researchers, and we used it mainly to explore the potential importance of the data for psychology (Whiten and Byrne 1988), just in case some or all of the records one day proved to be based on mental state attribution. We also discussed whether there was in fact any evidence as yet for that possibility, and noted only two cases that were difficult to account for more simply (Byrne and Whiten 1988b). Interesting as they were, it seemed unlikely at that time that these cases were any more than rogue observations, an inevitable risk of collecting in the indiscriminate way we had been forced to use. Nevertheless, some colleagues managed to read otherwise into our work, and criticised us for claiming rich intentional explanations for monkey behaviour, which we had not done. It may be that people, even scientists and colleagues, are biologically 'prepared' to over-attribute intentionality, even to the dry words of scientific papers (Byrne 1998).

The results of this informal survey justified a more systematic attempt, and in 1998 we enlisted the help of the International Primatological Society, the Primate Society of Great Britain, the Association for the Study of Animal Behaviour (in Europe) and the Animal Behaviour Society (in the USA). These organizations kindly allowed us to advertise in their newsletters, inviting scientists to submit any unpublished records, concerning primates or any other

animal species, that met a rather broad definition of tactical deception: 'acts from the normal repertoire of the agent, deployed such that another individual is likely to misinterpret what the acts signify, to the advantage of the agent'.

Note that this definition is a purely functional one: it does not require that the primate agent planned its tactics, intended to deceive another individual, realised that its tactic worked by deception; not does it imply that the victim was capable of realising that it had been deceived. Coming from a psychology background, we assumed that none of these things was at all likely. The central purpose of the survey was to chart the distribution of the trait in animals, and allow any patterns in this distribution to be revealed. However, it remained an open issue as to whether any evidence of intentionality might be forthcoming.

DISTRIBUTION OF TACTICAL DECEPTION

First, we analysed only those records that, even by the most stringent criteria, would fit our data. This approach was adopted in order to convince sceptical theorists, particularly in experimental psychology, that the phenomenon was genuine. However, it clearly risks the opposite, Type II error of disregarding perfectly genuine and valuable cases (recall that the original observers, rather than ourselves, were experts on their particular study species). Taking that risk, we required (Byrne and Whiten 1990) that:

- Usage was *unusual*, e.g. if an animal used a vocalisation known as a 'food call' to attract another when no food were present, that might be a case of tactical deception. But to be sure, we must know that the call normally functions *specifically* to indicate food. If, instead, it was a general attractor that had been too loosely labelled a 'food call', then there would be no need to conclude deception. Unless the definition of the call was rigorous, then, we discounted the record.
- Usage was *tactical*, e.g. it might be that when a subordinate grooms a dominant animal in possession of food, the dominant automatically relaxes so completely that it leaves hold of the food, and this allows the subordinate to grab it. In that case, the groomer gets food by grooming — but as a windfall, a lucky coincidence. Only if the same individual repeated the tactic, and specifically did so only when the groomee held food, could we be sure of more than coincidence.
- Modus operandi was *deception*, not some other form of social manipulation, e.g. grooming may regularly allow unrelated female monkeys access to a young baby, but as a 'trade' for the privilege of being groomed, rather than because the mother is misled about the plans of the groomer. So, whenever the agent gave something in exchange for its reward, we saw no need to invoke deception.

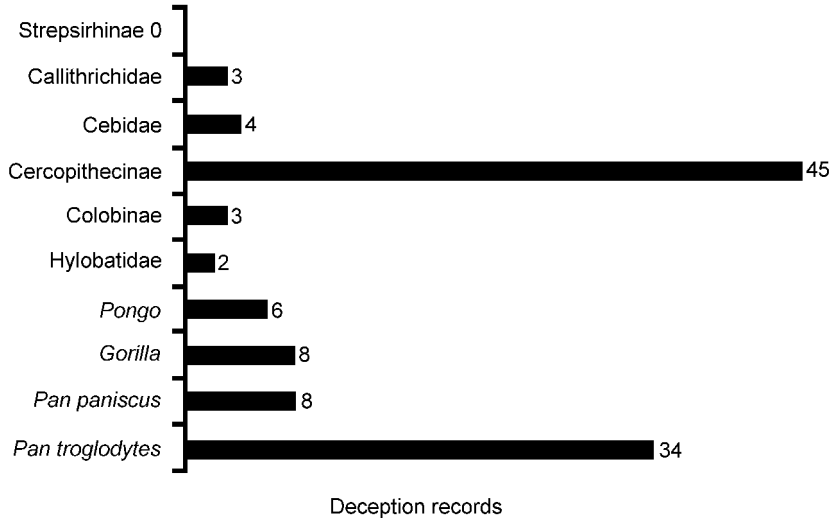


Figure 3.2. Distribution of tactical deception among primates

Having made this harsh sorting, we examined the distribution of what remained across the major taxa of primates (Figure 3.2). The only group missing entirely was the strepsirhine primates, the lorises and lemurs. Since most species are small and nocturnal, committing a Type II error was a particular risk in this group and, in fact, a few records of possible lemur deception had been submitted. The apparent strepsirhine/haplorhine difference in use of deceptive tactics may therefore be unsafe, and certainly all the main groups of haplorhines are represented by cases of tactical deception. This even applies to the colobines, a taxon which have often been dismissed as 'unintelligent' (the supposed justification for this attribution is that colobines are unintelligent *because* they are folivores, and can neither afford the metabolic costs of a large brain, nor do they have any need of the enhanced intelligence that enlarged brains would permit; Clutton-Brock and Harvey 1980).

The distribution of tactical deception was not random, however. What caused this variation among groups? Notice that, in Figure 3.2, the chimpanzee and the cercopithecine monkeys are over-represented. These species have been particularly well-studied in the wild; partly this is likely a function of their terrestrial ranging, which makes close observation easier. In addition, some species were historically thought to offer special insights into the anthropological reconstruction of early man, either because of their close relationship to humans (chimpanzee) or because of their similar ecological

niche as large-bodied, open-country primates (baboons, some species of macaques). The simplest explanation for the variation in the recorded incidence of tactical deception, then, would be that it reflects observer effort, rather than real differences among species. We tested this by comparing the distribution against the number of long-term field studies that might in principle have thrown up records, and the difference was significant (Byrne and Whiten 1992). Moreover, this effect was robust, remaining even when the species contributing most to the chi-square statistic was excluded (*Pan*, with few long-term studies but many records of deception).

The difference between the distribution of observer effort and the distribution of primate tactical deception must reflect some intrinsic feature of the primates themselves. Could it be a matter of intelligence? Animal intelligence cannot be reliably measured (Byrne 1996), so I used brain size as a proxy measure. Not long before, it had been discovered that the typical group size of primate species was correlated with the species' neocortex size (Dunbar 1992) and that, of the various possible measures of neocortex size, the neocortex ratio produced the strongest effects (the neocortex ratio is the ratio of the volume of the neocortex, compared to the volume of the rest of the brain). In an exploratory investigation, then, I also used the neocortex ratio (Byrne 1993). The frequency of deception records must partly depend on observer effort, so to produce a rough-and-ready index of the real frequency of deception, I divided the actual number of records by the number expected on the basis of observer effort alone. An index value greater than 1 indicates more deception than expected, whereas an index value below 1 indicates less than expected. To assay observer effort, I used the numbers of studies listed in a well-known study of the growth of primate field studies (Southwick and Smith 1986). Considering the 'noisiness' of a corpus of data contributed by hundreds of different scientists, it might be expected that no clear effect would emerge. It was therefore encouraging that a relationship *was* found: taxa with larger neocortex ratios were more often recorded as using tactics of deception in their everyday social lives, and this was a highly significant effect, $F_{1,8} = 11.9$, $p < 0.01$. Simply knowing the ratio of the brain volume taken up by the neocortex, divided by the volume of the rest of the brain, enables us to predict 60% of the variance in the amount of deception that is observed in the species concerned.

In that analysis, each data point was a monophyletic taxon (sometimes a single species, sometimes the average values of a larger group), so we may examine the details to see which taxa gave rise to the strong overall relationship between neocortex ratio and deception frequency (Figure 3.3). Notice that there is no clear separation between great apes and monkeys: some monkeys have larger neocortex ratios than some great apes, and some monkeys are more likely to be reported using deception than some great apes. Does this mean that, within the limit of error of this work, great apes and monkeys make

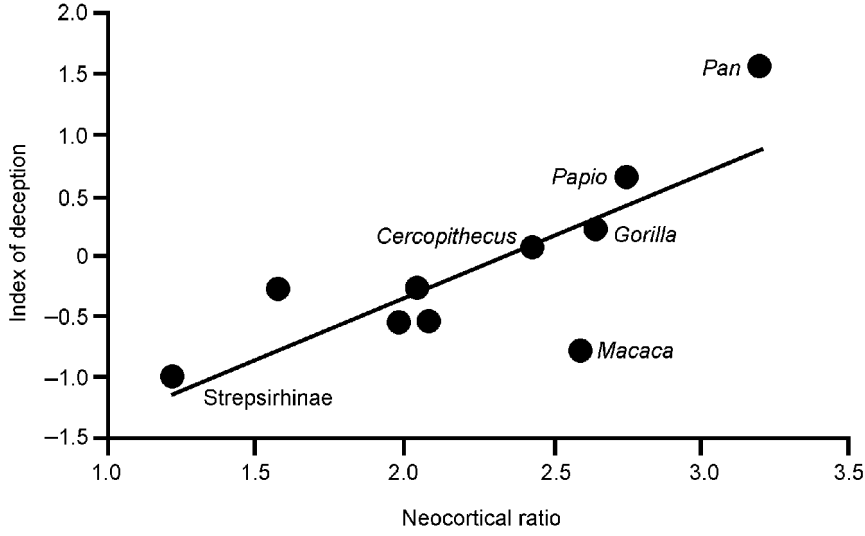


Figure 3.3. Correlation between observations of tactical deception and neocortex ratio. Some primate taxa are labelled

similar use of tactical deception, and the rate is affected only by their neocortex size? Not necessarily. Any one of the variables — the index of deception, the measure of neocortex size, and the method of analysis — might have led to error. Moreover, my exploratory analysis did not consider the possibility of artifact from a correlated variable. Therefore, to discover whether this effect was reality or chimera, Nadia Corp and I have recently repeated it, dealing — we hope — with the possible problems.

First, consider the deception index. The records of tactical deception used were those rated in the published corpus as assuredly being cases of tactical deception (level 1 and above; for details, see Byrne and Whiten 1990). This was appropriate for convincing sceptics of the reality of the phenomenon, but perhaps not for making a fair assessment of frequency. Systematic bias might have been introduced if Whiten and I had some unconscious preconceptions about which narratives were the most convincing evidence of deceptive tactics. Could we have been more lax in our standards for some species than others? Corp and I therefore returned to the raw data, the records as originally submitted by the primatologists themselves. At the same time, we improved accuracy of method of scaling for observer effort. Southwick and Smith's analysis of field studies described a slightly earlier period of field study (pre-1986), and the balance among species might have shifted subtly by the time (post-1988) when informants were contributing to the 1990 corpus. Instead, we

used the data from *Current Primate Field Studies* (regularly published by the Primate Society of Great Britain) for the precise years over which the corpus was largely assembled. All studies that were long-term, i.e. more than 2 years, and whose focus was not entirely survey or conservation work, were counted.

Use of the neocortex ratio is also problematic, since it is not independent of brain size (Harvey and Pagel 1991). Moreover, on average, brain size and body size are closely correlated, which means that the neocortex ratio will also vary with body size. Other things being equal, a larger animal has a larger neocortex ratio. Could it be that large primates are more often reported to use deception, perhaps because subtle tactics are easier to see in larger subjects? To avoid any such concerns, and obtain a measure that is independent of brain and body volume, we first regressed neocortex volume upon total brain volume. We then took the *residuals* from the best-fitting line to indicate the extent to which the neocortex is larger or smaller than expected from brain size alone: these are independent of both brain and body size (but see below for a further important complexity).

Since group size is known to correlate with neocortex ratio, it might be that the significant relationship between neocortex size and deception frequency was mediated by group size. When primatologists study a species typically living in large groups, they inevitably experience more opportunities to notice social manipulation. Also, the primate subjects themselves might plausibly experience more social situations needing subtle tactics if they live in larger groups. Either way, an artifactual relationship with neocortex is possible, as a result of a primary relationship between group size and deception. To check for this possibility, we included the species' typical group sizes in our analysis.

My original analysis used taxonomic groups as data points; however, each taxon is more or less closely related to each other, so the points are not statistically independent. This is a problem long-recognized in scaling studies, and I had used a traditional remedy — using higher taxonomic groupings (genera) instead of individual species. At the genus level, the problem is somewhat reduced; however, it is not eliminated. Also, averaging several separate species may actually be throwing away meaningful variation. Much better is to use the method of independent contrasts (Harvey and Pagel 1991; Pagel 1992) to eliminate taxonomic bias entirely. This approach begins from a cladogram that represents best the evolutionary relationships among all species in the group, as a series of binary divisions. Each branch point represents a hypothetical evolutionary event, and the variables of interest are compared between the two branches (the value for each branch is computed from its subsidiary branches). With n species for which one has data, $n-1$ independent contrasts are obtained, and these contrasts are then examined for relationships. In this case, we had 17 species for which data on deception and brain parts were available, spanning the entire primate radiation: three prosimians, four New World monkeys, seven Old World monkeys (five cercopithecine and two

colobine), and three ape species. For this set of species, we extracted phylogenetic contrasts in the tactical deception index, in species group size, and in both neocortex volume and total brain volume (16 in each case). Regressing the contrasts in neocortex volume against those for total brain volume gave relative contrasts in neocortex volume, our measure of neocortical investment. These variables were then entered in a multiple regression, using group size and relative neocortex size to explain variation in tactical deception. The regression was performed stepwise, and also with group size forced first into the predictive equation.

Whichever way we performed the regressions, the same effect emerged: relative neocortex size strongly predicts use of tactical deception (Byrne and Corp in preparation). Intriguingly, group size did not. In stepwise regression, this variable was not selected. Even when it was forced into the regression first, it was not a significant predictor of deception frequency. One point was somewhat of an outlier, the contrast between *Pan* and *Gorilla*; however, the strong and significant relationship remained when it was excluded. We tentatively interpret the outlier to mean that, although closely related, the chimpanzee and gorilla lines have diverged recently. In their investment in relative brain parts, the chimpanzee specializes more in neocortex and shows more social deception as a consequence, and the gorilla invests more in other brain areas, perhaps the cerebellum (which is particularly large in *Gorilla*; although note that, in absolute terms, the gorilla also has a larger neocortex than *Pan*).

I therefore conclude that neocortical enlargement, which is prominent in the haplorhine primates as a whole, allows greater use of complex means of social manipulation, such as tactical deception. Since the analysis was based on a functional definition, this is by no means to imply that these species have any understanding of how the tactics work, or that they might plan their deceits. Instead, rapid learning would be capable of underwriting these tactics, in species which are capable of remembering the identities and past actions of a number of social companions, and sensitive to those individuals' presence or nearness in the social milieu (Byrne 2000). The simian clade (monkeys and apes) has undergone selection for rapid learning in social contexts, resulting in relatively enlarged neocortex and greater reliance on social manipulation within the group, especially in those species with the most enlarged neocortex.

SIGNS OF UNDERSTANDING IN PRIMATE TACTICAL DECEPTION

Although Whiten and I did not particularly expect to find evidence of insight into how the tactics worked, we did find a little. In a number of cases, features of the records were very hard to account for, without allowing that the primates had the cognitive capacity to envisage how the situation would look

from the target's point of view. The approach we took was to examine each record for whether either of us could devise an associative conditioning explanation that was remotely plausible, and then separated out the cases where it was not. The most well-known record of this kind is an observation of Hans Kummer. It concerns a hamadryas baboon female who was noticed to spend 20 min in shifting her body, millimetre by millimetre, until she was in a position to groom a lower-ranking male, but not be seen to do so by her harem-leader male. This is sometimes represented in cartoon form, in a way that implies theory of mind (Figure 3.4a): 'I believe that he thinks there is no other baboon near me'. This is unfortunate, because a simpler account is also possible, in which the female is able to compute the view from another person's perspective, and what it means to them (Figure 3.4b): 'I believe he cannot see behind my rock from his position', or, even simpler (Figure 3.4c); 'If I were sitting where he is, I could not see what is behind this rock'. In our analysis in both 1990 and 1992, we treated cases like this as a separate category, not entailing any theory of mind on the part of the primates; we labelled the category *visual perspective-taking*. A few records went beyond purely visual perspective-taking and strongly suggested that the animals had knowledge about what others knew or thought. No one of these records, taken alone, would be entirely compelling—just as, perhaps, no experimental demonstration of theory of mind, if it conflicted with all other data, would be acceptable. Indeed, we showed that each one could be modelled as the output of a set of mechanical rules ('productions', in the jargon of computer simulation), and that each single rule is, in principle, learnable (Byrne and Whiten 1991). However, the circumstances which would have had to happen, by coincidence, in order to condition these rules, were sometimes rare to the point of approaching the bizarre. For 18 cases, Whiten and I both independently rated it as more parsimonious to invoke some simple understanding of mental states on the part of the primates than to posit the strange imaginary histories that a conditioning explanation required. We treated these records as evidence of *mental perspective-taking*.

Visual and mental perspective-taking had strikingly different distributions among the corpus of records, although neither was common. Visual perspective-taking was scattered across the taxonomic groupings, in rough proportion to the abundance of data submitted. At least in the well-studied cercopithecine monkeys and the common chimpanzee and other great apes, some ability to compute information about the viewpoint of another individual and what it can and cannot see appears to be general. In striking contrast, mental perspective-taking was sharply concentrated on the non-human great apes (Byrne and Whiten 1992). Unlike the case of visual perspective-taking, monkeys did not provide any convincing evidence at all that they could take the mental perspective of another individual. That, of course, is not in itself surprising: in 1992, and since, monkeys have provided no convincing evidence

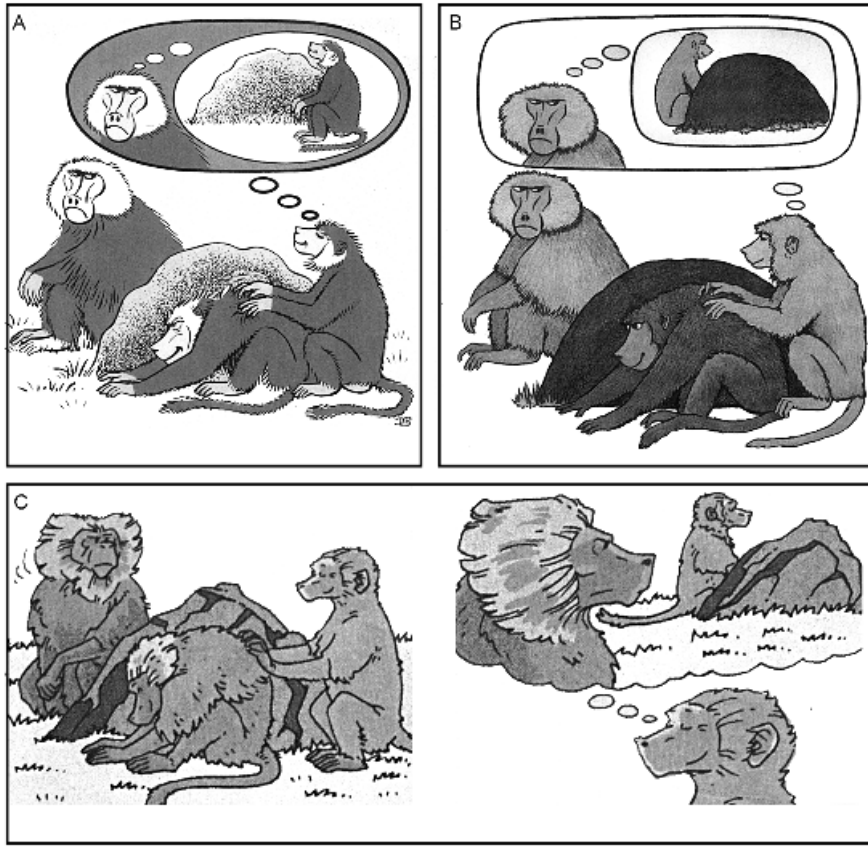


Figure 3.4. Differing interpretations of deception (see text for details of the Kummer observations to which these cartoons refer): (a) attribution of false belief; (b) attribution of ignorance; (c) attribution of visual perspective

of theory of mind in experiment or field observations (Tomasello and Call 1997). What was more surprising (in 1992) was that great apes should provide such evidence and that all species were represented, not simply the common chimpanzee. In 2000, there is at last some convincing work coming out of laboratories to support this picture, e.g. Boysen (unpublished) has demonstrated that chimpanzees are sensitive to whether their audience knows of an approaching danger or not; and Call and Tomasello (1998) found that chimpanzees and an orangutan were able, immediately, without any history of differential reward, to distinguish between accidental and deliberate marking of a box that might contain food. Clearly, these results need replication and

extension, but it may tentatively be expected that some level of mental perspective-taking will prove to be general in the great apes.

This visual/mental distinction in perspective-taking has been largely ignored, and most attempts to show experimentally that the common chimpanzee is able to appreciate the visual perspective of others have seemingly believed that this would be tantamount to having a theory of mind (e.g. Povinelli, Nelson and Boysen 1990). However, Hare *et al.* (2000) have used a similar distinction in accounting for their (positive) data on visual perspective-taking in the chimpanzee. The challenge for experimenters now is to detect this visual perspective-taking ability in monkeys, as predicted by the distribution of the observational data on tactical deception.

IMPLICATIONS FOR THE EVOLUTION OF HUMAN COGNITION

Deception is but one of many areas of primate cognition under active investigation, and even the results sketched in this chapter are not accepted as definitive by all researchers. It is therefore merely to illustrate the practice and potential of evolutionary reconstruction in understanding cognitive evolution, not with any hope of putting forward the last word on the origins of the human mind, that I will go on to sketch the implications of accepting these data.

First, consider the distribution of tactical deception, as illustrative of a sophistication in social manipulation beyond what is seen in all mammals, and neocortical enlargement, as illustrative of brain specialization relating to social complexity. Both these characters are more notable in the simian primates, monkeys and apes, compared with strepsirhine primates and most non-primate mammals. That implies an origin in the common ancestor of all monkeys and apes, at about 30 million years ago, which was presumably more highly social than its ancestors (Byrne 2000). Most plausibly, the spur to evolutionary investment in metabolically costly neocortical tissue was an increase in the group size of the ancestral simian, itself a consequence of increased predation pressure because these species were daytime-living, whereas their ancestors were nocturnal. These findings support the 'Machiavellian intelligence' or 'social brain' hypothesis of human intelligence, as far as aspects of intelligence we share with all simians (Jolly 1966; Humphrey 1976; Byrne and Whiten 1988a). Note, however, the limited nature of these specializations. The amount of tactical deception used in social living varies quantitatively with a species' neocortical size; the incidence does not drop to zero outside simian primates (Hauser 1997), although the low frequencies may often result in dismissal as 'anecdotal' in other mammals. Moreover, there is no evidence that simians in general have any understanding of the mechanism of deception, i.e. the creation of false beliefs (Tomasello and Call 1997). Instead, rapid learning in

social contexts has been argued to underwrite the use of deception and other complex social tactics (Byrne and Whiten 1997). Neocortex size, then, may relate to efficiency of learning about social parameters, including an appreciation of the spatial organization of the group and line-of-sight relationships among other individuals, but no particularly deep understanding of what any of these parameters mean.

Turning to the origins of more insightful intelligence, going beyond merely visual perspective-taking to have some understanding of how other people work, then an origin in the common ancestor of all the great apes is implied — because it is only in the great apes that there is any evidence of such abilities. This qualitative difference in understanding, however far removed from the social insight of modern humans, marks the first glimmerings of the remarkable suite of abilities we now label 'theory of mind', dating to about 12 million years ago. Its origin cannot be attributed to social complexity, since modern great apes are not systematically more socially complex than monkeys: some monkey species live in larger, more stratified societies than any ape, without any special signs of insightful cognition, and some apes are close to solitary. It may be argued that insightful understanding is an emergent property of brain size *per se*, as all modern great apes have absolutely larger brains than any monkey, since they are very much larger animals. However, this does not account for the ultimate cause, large body size. Alternatively, it may be that insightful understanding comes from a difference in mental representation, a 'software' adaptation, in response to pressures to feed more efficiently in competition with sympatric monkeys (Byrne 1997).

POSTSCRIPT: HOW SHOULD OBSERVATIONAL DATA BE USED?

I hope it is clear from this account that the commonly portrayed role of observations, of no more than pointing to possibilities that can then be tested rigorously by experiment, is misleading. Observations must be meticulously collected by experienced observers: standards of scientific observation need to be *higher*, not lower, than in experimental work. Conversely, experiments should build on the naturally adapted capacities of animals: experiments that treat animals as models of inferior humans, as if retarded or damaged, are unlikely to generate meaningful data. With those provisos, both observation and experiment can give solid data. Observations need careful interpretation, and interpretations may change over time; but these limitations apply with equal force to experimental results. Each has its advantages. Assembling proper control groups is much harder from spontaneous observation than it is in an experiment. Experiments will always tend to focus on the few species widely available in captivity, or particularly easy to work with in the wild;

observations can fill in the broader picture, so essential for evolutionary reconstruction. As is clear in the case of deception, observations are often in the lead: this does not necessarily imply that the observers are weak minded and desperate to believe their animals are nearly human. Ideally, both methods should converge on a consistent picture of cognition. Reality will no doubt often fall short of this ideal, but a doctrinaire over-reliance on experimentation will only impede science.

REFERENCES

Byrne, R.W. (1993). Do larger brains mean greater intelligence? *Behav. Brain Sci.* **16**, 696–697.

Byrne, R.W. (1996). Relating brain size to intelligence in primates. In P.A. Mellars and K.R. Gibson (eds), *Modelling the Early Human Mind*. Macdonald Institute for Archaeological Research, Cambridge, pp. 49–56.

Byrne, R.W. (1997). The technical intelligence hypothesis: an additional evolutionary stimulus to intelligence? In A. Whiten and R.W. Byrne (eds), *Machiavellian Intelligence II: Extensions and Evaluations*. Cambridge University Press, Cambridge, pp. 289–311.

Byrne, R.W. (1998). So much easier to attack straw men. *Behav. Brain Sci.* **21**, 116–117.

Byrne, R.W. (2000). The evolution of primate cognition. *Cogn. Sci.* **24**, 543–570.

Byrne, R.W. and Whiten, A. (1985). Tactical deception of familiar individuals in baboons (*Papio ursinus*). *Anim. Behav.* **33**, 669–673.

Byrne, R.W. and Whiten, A. (1988a). *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans*. Clarendon, Oxford.

Byrne, R.W. and Whiten, A. (1988b). Towards the next generation in data quality: a new survey of primate tactical deception. *Behav. Brain Sci.* **11**, 267–273.

Byrne, R.W. and Whiten, A. (1990). Tactical deception in primates: the 1990 database. *Primate Rep.* **27**, 1–101.

Byrne, R.W. and Whiten, A. (1991). Computation and mindreading in primate tactical deception. In A. Whiten (ed.), *Natural Theories of Mind*. Basil Blackwell, Oxford, pp. 127–141.

Byrne, R.W. and Whiten, A. (1997). Machiavellian intelligence. In A. Whiten and R.W. Byrne (eds), *Machiavellian Intelligence II: Extensions and Evaluations*, Cambridge University Press, Cambridge, pp. 1–23.

Byrne, R.W. and Whiten, A.W. (1992). Cognitive evolution in primates: evidence from tactical deception. *Man* **27**, 609–627.

Call, J. and Tomasello, M. (1998). Distinguishing intentional from accidental actions in orangutans (*Pongo pygmaeus*), chimpanzees (*Pan troglodytes*), and human children (*Homo sapiens*). *J. Comp. Psychol.* **112**, 192–206.

Clutton-Brock, T.H. and Harvey, P. H. (1980). Primates, brains and ecology. *J. Zool. Lond.* **190**, 309–323.

Corbet, G.B. and Hill, J.E. (1991). *A World List of Mammalian Species*, 3rd edn. Oxford University Press, Oxford.

Dunbar, R.I.M. (1992). Neocortex size as a constraint on group size in primates. *J. Hum. Evol.* **20**, 469–493.

Goodall, J. (1971). *In the Shadow of Man*. Collins, London.

Hare, B., Call, J., Agnetta, B. and Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Anim. Behav.* **59**, 771–785.

Harvey, P.H. and Pagel, M. (1991). *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford.

Hauser, M.D. (1997). Minding the behaviour of deception. In A. Whiten and R.W. Byrne (eds), *Machiavellian Intelligence II: Extensions and Evaluations*. Cambridge University Press, Cambridge, pp. 102–143.

Humphrey, N.K. (1976). The social function of intellect. In P.P.G. Bateson and R.A. Hinde (eds), *Growing Points in Ethology*. Cambridge University Press, Cambridge, pp. 303–317.

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

Pagel, M.D. (1992). A method for the analysis of comparative data. *J. Theoret. Biol.* **156**, 434–442.

Povinelli, D.J., Nelson, K.E. and Boysen, S.T. (1990). Inferences about guessing and knowing by chimpanzees (*Pan troglodytes*). *J. Comp. Psychol.* **104**, 203–210.

Southwick, C.H. and Smith, R.B. (1986). The growth of primate field studies. In G. Mitchell and J. Erwin (eds), *Comparative Primate Biology. 2a. Behaviour, Conservation and Ecology*. Liss, New York.

Tomasello, M. and Call, J. (1997). *Primate Cognition*. Oxford University Press, New York.

Whiten, A. and Byrne, R.W. (1988). Tactical deception in primates. *Behav. Brain Sci.* **11**, 233–273.

4

ProtocadherinXY — a Candidate Gene for Cerebral Asymmetry and Language*

TIM J. CROW
Warneford Hospital, University Department of Psychiatry, Oxford, UK

THE PROBLEM FOR EVOLUTIONARY THEORY

'Human language is an embarrassment for evolutionary theory', perhaps, as Premack (1985, p. 283) suggested, 'because it is vastly more powerful than one can account for in terms of selective fitness', but also perhaps as Chomsky (1972) had pointed out, in language we have a faculty without clear precedent in other primates (see also Penner 2000). Bickerton (1995) has argued that language appeared suddenly and recently. Evidence such as rock art for a representational capacity that might parallel possession of language can be traced back, with some difficulty, to around 90,000 years ago (Noble and Davidson 1996; Mellars 1998). Such a capacity is absent in the archaeological record that relates to the Neanderthals (Mellars 1998), as also in the longer record that relates to *Homo erectus* (Bickerton 1995; Noble and Davidson 1996). While these hominid species are credited with 'proto-language', the components of language (e.g. grammatical elements and structure, embedding and subcategorization of verbs) characteristic of modern humans are assumed to have been absent. Thus, the acquisition of the capacity for language appears

*Reprinted from A. Wray (ed). (2002). *The Transition to Language*. Oxford University Press, Oxford.

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

to represent a recent, and datable, discontinuity in hominid evolution. The parsimonious conclusion (because it links the distinctive characteristic of the species to its genetic origin) is that the origin of language coincided with the transition to modern *Homo sapiens*, dated to somewhere between 100,000 and 150,000 years ago (Stringer and McKie 1996).

This conclusion forces us to take sides in a long-running dispute about the nature of the evolutionary process, particularly as it relates to speciation. The focus of the debate is on what genetic change accounts for the transition from one species to another and by what selective process it is retained.

GRADUALISM IN EVOLUTION

The prevailing view within evolutionary theory — the biological or isolation species concept — is that populations separate, become subject to different environmental selective pressures, and acquire genetic variation that leads to reproductive isolation (Dobzhansky 1937; Mayr 1963). This view is consistent with Darwin's original concept, represented, for example, in the only figure that appears (in Chapter 4) in the *Origin of Species* (Darwin 1859). The figure indicates that over the passage of thousands of generations, variations accumulate and lineages separate. Some lineages are extinguished, and those that survive may be widely separated in geography and perhaps also in adaptive characteristics. Darwin was anxious to emphasise the continuity of the variation within and between species. No qualitative distinction was drawn.

However, in the case of the origin of language, viewed as a speciation event, two problems arise with this view. The first is that there is no evidence of a gradual accumulation of linguistic capabilities over a long period. The second is that the dispersal and geographical isolation of modern humans must have occurred *after* the appearance of language, with the propensity for language being present in all humans *despite* that dispersal and isolation. The alternative would be to assume that such changes occurred independently but with the same effect in populations that had already separated. To suggest this is to invoke an implausible process of parallel evolution in separate continents. Clearly the genetic events being considered here, as also those that define any species, must be limited in time. But *how* restricted in time and space a speciation event may be expected to be is the subject of debate within evolutionary theory (see e.g. Otte and Endler 1989; Howard and Berlocher 1998; Coyne and Orr 1998; Magurran and May 1999).

DISCONTINUITY IN EVOLUTION

Discontinuity theory thus challenges Darwinian gradualism. Amongst the earliest proponents of discontinuity were Bateson (1894) and, in the wake of

the re-discovery of Mendelian principles, de Vries (1905). A radical attack was mounted by Goldschmidt (1940), with his concept of 'hopeful monsters', the outcome of macromutations that generated innovations in a single step. With reference to the time course of evolutionary change in the fossil record, Eldredge and Gould (1972) reintroduced the question of discontinuity with their concept of 'punctuated equilibria'.

Theories of discontinuity face two general problems of their own, however:

1. How is sudden change selected? If change is not gradual as Darwin supposed, by what new principle does speciation occur? The problem can be illustrated with reference to Goldschmidt's 'hopeful monsters'. How can a radical departure from a body plan previously adjusted by a long process of environmental selection be an adaptational improvement?
2. By what genetic mechanism are 'between-species' variations to be distinguished from 'within-species' variations?

One answer (Goldschmidt's solution) to the second question is that the critical changes that distinguish one species from another relate particularly to chromosomal change, i.e. to structural rearrangements of the chromosomal complement. These would affect the capacity for successful reproduction between individuals having the different arrangements, by leading to infertility or sterility of the offspring. While this mechanism has been strongly promoted by some authors (e.g. White 1973; King 1993), others (e.g. Coyne and Orr 1998) are unconvinced of any necessary relation between chromosomal change and species transitions. For example, it is argued that inter-species hybrid sterility may be present when no structural changes are detectable, and that some within-species variations in chromosomal structure are unassociated with a reduction in fertility. Where obvious structural differences between species are present, there is the possibility that these accumulated after, rather than at the time of, the speciation event.

In this chapter I shall suggest (following Crow 1993a,b, 1996, 1998a,b, 2000) that consideration of language and its relation to the speciation of modern *Homo sapiens* offers a possible solution to these problems, specifically that it is a subset of chromosomal changes — those relating to the sex chromosomes — that has particular relevance to species transitions, and that this is the case because such changes (perhaps particularly those relating to regions of homology, i.e. DNA sequence similarity, between the two sex chromosomes) can be subject to a process of sexual selection, a process that refines and adapts the consequences of the primary change. Sexual selection is the process of mate choice that Darwin distinguished from natural selection to explain features such as deers' antlers and the peacock's tail that differentiate one sex from the other within a species. The sequence of events in which a chromosomal change is followed by sexual selection is consistent with the implication of Darwin's (1871) juxtaposition of his treatise on the descent of man with his theory of

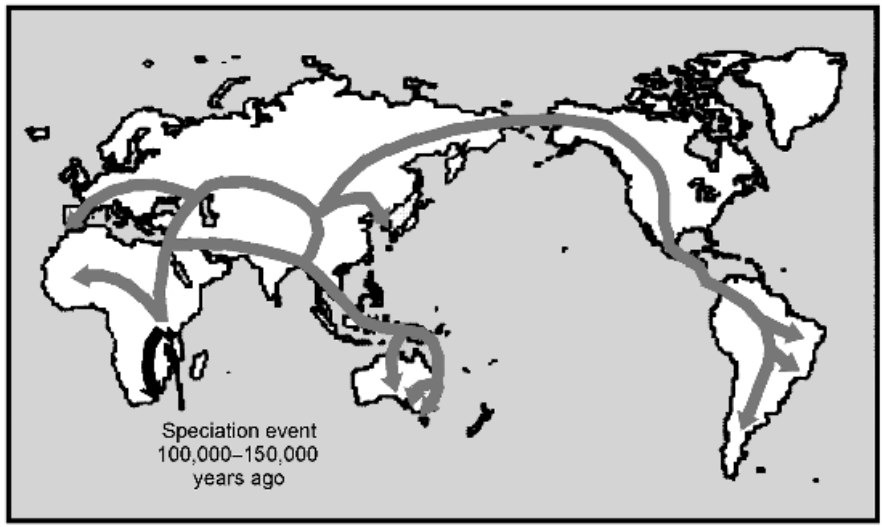


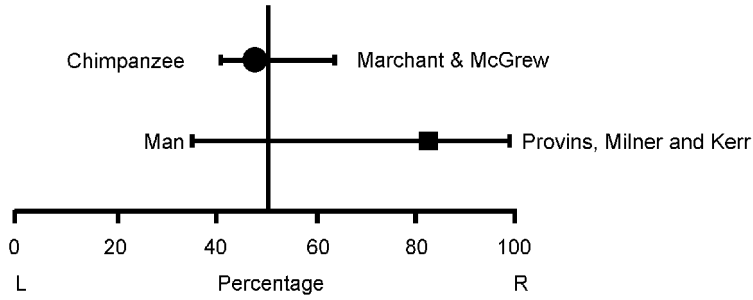
Figure 4.1. The diaspora of modern *Homo sapiens* (adapted from Stringer and McKie 1996, p. 169) to emphasise the significance of the genetic change (designated a speciation event) that enabled the transition from a prior hominid species

sexual selection — that *Homo sapiens* had evolved by some process of sexual selection, although this implication seems not to have been made explicit.

THE OUT-OF-AFRICA HYPOTHESIS

The current context of discussions of human evolution is the Out-of-Africa hypothesis — the theory that modern *Homo sapiens* originated some time between 100,000 and 150,000 years ago as a result of a genetic change that occurred in a population somewhere in East Africa (Stringer and McKie 1996; see Figure 4.1). A parsimonious view is that it was this change that accounted for the transition from 'proto-language' to full human language and that it is the latter capacity that accounts for the extraordinary biological success of *Homo sapiens* compared to precursor primate and hominid species. That language is the defining feature of humanity seems first to have been clearly stated by de Condillac (1746). The universality of the capacity for language to human populations (as appreciated by Sapir 1921) reflects the genetic identity of the species, an identity which is invariant with respect to the environment.

But what could that change have been? There are few candidates, but the one that has been available at least since the observations of Broca 135 years ago, is



	Number of Individuals	Activities
Chimpanzee	38	46
Man	1960	75

Figure 4.2. Directional handedness in *Pan troglodytes* and *Homo sapiens*. Marchant and McGrew (1996) assessed hand usage in 38 chimpanzees observed in the Gombe National Park in Kenya and Provins, Milner and Kerr (1982) made similar assessments of the range of everyday activities in a human population (934 individuals). The bar graph indicates the percentage of activities engaged in performed by each hand

that the brain lateralised, and that some component of language was confined to the dominant hemisphere. The singularity of brain lateralisation is strongly supported by cross-species comparison within primates. Handedness, the outward manifestation of hemispheric dominance, is strongly skewed to the right in human populations (Provins, Milner and Kerr 1982; Perelle and Ehrman 1994) and Annett (1985) has argued that it can be accounted for by a single gene. In contrast, according to the observations of Marchant and McGrew (1996) of 38 chimpanzees in the wild in the Gombe National Park, while hand preference to the right or to the left may be a characteristic of the individual directional bias, at the level of the population it is absent. No other index so clearly distinguishes the two species (Figure 4.2).

It appears that directional handedness is an outward manifestation of an anatomical asymmetry in the human brain that is probably best represented as a 'torque' (Figure 4.3). In most individuals the right frontal lobe is wider than the left and the left occipital lobe is wider than the right. Frontal and occipital asymmetries are intercorrelated (Bear *et al.* 1986) and are reflected in the asymmetry of the planum temporale, described by Geschwind and Levitsky (1968). A correlate at the cellular level has been uncovered by Buxhoeveden and Casanova (2000). They report that the mean distance between the columns of pyramidal cells in the superior temporal gyrus (Wernicke's area),

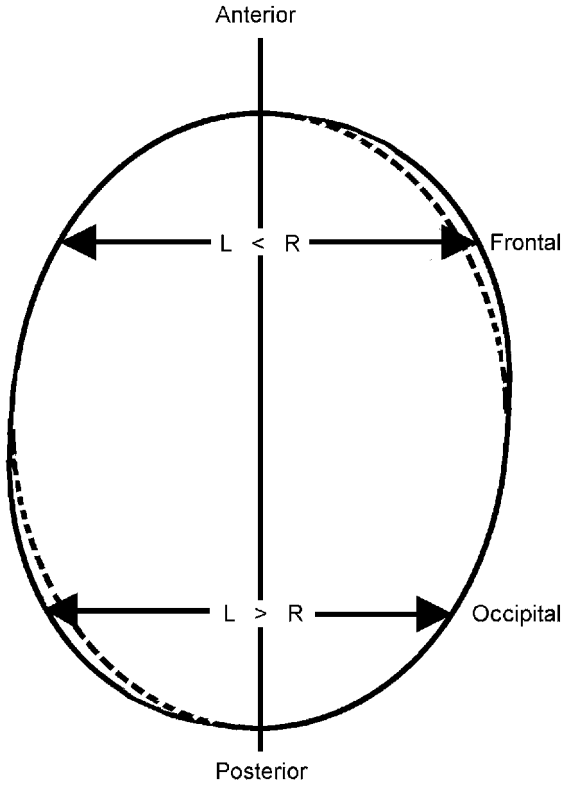


Figure 4.3. The anatomical 'torque' in the human brain from right frontal to left occipital

asymmetrical to the left in the human brain, does not differ between the hemispheres in the chimpanzee.

SEX DIFFERENCES AND X–Y LINKAGE

Sex has a major influence on the development of language (Maccoby and Jacklin 1975; Halpern 1992). Girls talk at an earlier age than boys (Moore 1967) and have greater verbal ability (Butler 1984). What explains this sex difference? One possibility is that lateralisation for language is related to sex. In an examination of the UK National Child Development cohort, it was found that at the age of 11 years, on an index of relative hand skill, girls were more strongly lateralised than boys. Furthermore, degree of lateralisation predicted verbal ability (Crow *et al.* 1998). Those close to the point of 'hemispheric indecision' (equal hand skill) were impaired in the acquisition of

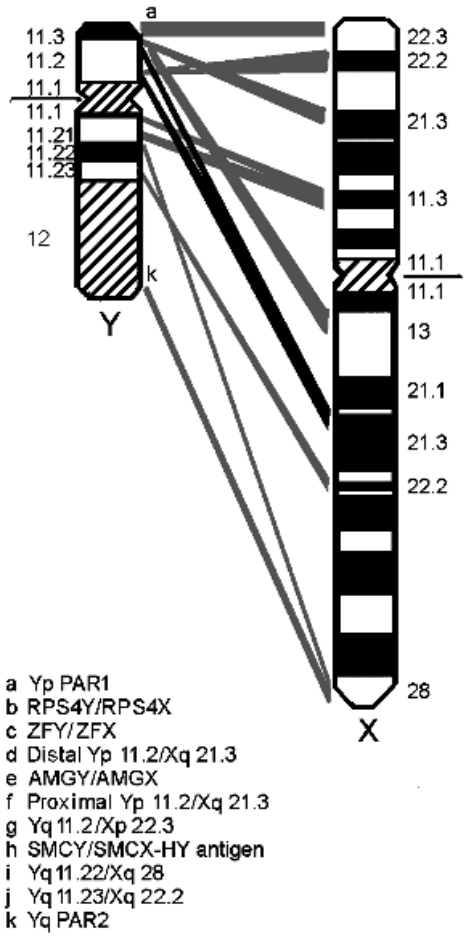


Figure 4.4. Diagram indicating regions of homology between the X and the Y chromosome. Bands of homology are labelled a–k on the Y chromosome. The dark band indicates the Xq21.3 region that is homologous to two blocks (d and f) on Yp (the Y short arm); k indicates the long arm (Yq) pseudoautosomal region (YqPAR2). Both these regions of homology are present in man but not in earlier primates. The short arm (YpPAR1) pseudoautosomal region was created earlier in evolution, as also were the other regions of homology. The region of the centromere is cross-hatched. Genes within established homologues on X and Y are identified by acronyms (RPS4X/Y, ZFX/Y, AMGX/Y, SMCX/Y). Adapted from Affara *et al.* (1996)

words relative to the rest of the population. Girls at this age had significantly more words than boys but the relationship of verbal ability to hand skill was the same in the two sexes. Thus, the acquisition of words may reflect the rate at which dominance is established in one hemisphere, girls being more

lateralised or lateralising faster than boys (Shucard, Schucard and Thomas 1987). It is of interest that, as far as is known, this difference in the acquisition of words is not accompanied by a sex difference in grammatical ability.

A clue to the genetic mechanism comes from observations of individuals with sex chromosome aneuploidies (a deficit or an excess of sex chromosomes relative to the normal two — XX in a female and XY in a male). Individuals who lack an X chromosome (Turner's syndrome) have relative impairments of non-dominant hemisphere ability, whilst those who have an extra X (XXY, Klinefelter's or XXX syndromes) have dominant hemisphere deficits. This strongly suggests that an asymmetry determinant [the 'right shift factor' to use Annett's (1985) term for the cerebral dominance gene] is present on the X. However, the fact that normal males (XY) do not have deficits in non-dominant, i.e. spatial, ability comparable to those that are seen in Turner's syndrome, indicates that the presence of the Y chromosome must complement the influence of a single X chromosome, i.e. the gene must also be present on the Y. Thus, it can be concluded that the gene is in the relatively select class that is present in homologous form on both the X and the Y chromosome (Crow 1993b; see Netley 1998). Evidence consistent with X–Y linkage was obtained from a family study (Corballis *et al*. 1996); X linkage is supported by a recent analysis of the literature and a new family collection (McKeever 2000).

X–Y homologous genes are an unusual class, most being generated by translocations of blocks of sequences on the X to the Y chromosome, i.e., there is a duplication on the Y of sequences that were previously present only on the X (see Figure 4.4). These translocations can be dated in the course of mammalian evolution (Lambson *et al*. 1992). Those of greatest interest are the two events that have occurred since the separation of the chimpanzee and hominid lineages:

1. A translocation from the X chromosome long arm (the Xq21.3 region) to the Y chromosome short arm (Yp) that is estimated to have taken place 2–3 million years ago (Sargent *et al*. 1996). The translocated segment on Yp was split by a subsequent paracentric inversion that left two blocks in Yp that are homologous to the original block in Xq21.3. The paracentric inversion has not been dated (Mumm *et al*. 1997; Schwartz *et al*. 1998), although this is a question of great interest (see Laval *et al*. 1998 for evidence of linkage of handedness to the Xq21.3 region).
2. The generation of the second pseudoautosomal region at the telomeres (ends) of the long arms of the X and Y chromosomes (Freije *et al*. 1992).

The Xq21.3/Yp translocation (as probably also the events relating to the second pseudoautosomal region) created new Y-linked representations of X-linked genes and thereby generated a balance of control of these genes between the sexes that was not previously present. Such an event occurred in a single male. The immediate effect was to double the dose of the relevant genes in that

individual (relative to other males, and to females, in whom the genes on one of the two X chromosomes are normally 'inactivated', a presumed mechanism of 'dosage compensation' between males and females). For such a translocation to become universal in the human population, the characteristic coded for by the gene clearly must have been selected. Because the new or accentuated characteristic was present in males, this selection may have been sexual, that is, males with the characteristic would have been preferentially selected by females as mates.

But since the gene was already present on the X chromosome, males now expressing the gene in double dose may, in turn, have subjected the relevant characteristic in females to a new selective force. Thus, one can envisage an escalating process whereby a characteristic first selected in males was then selected in females. The quantitative expression of the trait would differ between the sexes, as a function of two variables: (a) variation in the extent of inactivation of the copy of the gene on the inactivated X chromosome in females; and (b) variation in the gene sequence on the X and the Y chromosomes (see below, in relation to protocadherinXY). Thus, the disruption of a previous equilibrium by a single X–Y translocation (a discrete 'saltational' change) has the potential to have generated a process of sexual selection that led to the progressive refinement of a single feature to define a new species.

In the case of lateralisation, the initial step towards either 'proto-language' at an earlier stage, or towards full language in the case of modern *Homo sapiens*, could be honed by mate selection on a simple parameter, such as the point of maturation (Crow 1998a; see Figures 4.5 and 4.6).

One can raise the question of why, if two doses of the gene are advantageous after the translocation from the X to the Y, such adjustment could not have taken place when the gene was present only on the X. Because genes on one X chromosome in females are subject to inactivation (the dosage compensation mechanism referred to above), the gene sequence and the expression of the gene are the same in both sexes. Only after the gene has been reduplicated on the Y chromosome does the possibility of differential expression in males and females, and the potential influence of mate choice (sexual selection), arise. It seems possible that it is the process of sexual selection, acting at least in part by modifying the state of dosage compensation of genes on the X, that establishes the new evolutionary equilibrium. The difficulty in understanding the mechanism is that it is presently unclear how (in what appears to be a general rule for mammals) genes on the X with a homologue on the Y become protected from the process of X inactivation in females. One proposal is that pairing of X and Y sequences in male meiosis plays a role (Crow 1991). Investigating the phenomenon of protection from X inactivation across mammalian orders, Jegalian and Page (1998) suggest that, following the establishment of the sequence on the Y chromosome, change in gene sequence

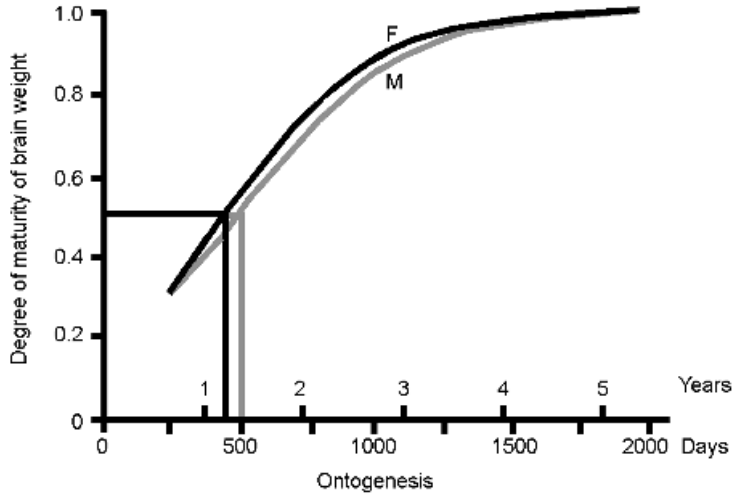


Figure 4.5. The sex difference in the rate of brain growth according to the data of Marchand. Reproduced from Kretschmann *et al.* (1979)

and inactivation status occurs by selective pressure, first on the male and then on the female, consistent with a role for sexual selection. The proposal, therefore, is that a gene within the Xq21.3/Yp11.2 region of homology (protocadherinXY is the specific candidate) has been subject to sexual selection in hominids, including modern *Homo sapiens*, because it is present on the Y chromosome as well as the X, whereas in the great apes it is present only on the X, and that changes in gene sequence on the Y and inactivation status on the X have been intimately involved in this process.

SEXUAL SELECTION AND SPECIATION

A relationship between sexual selection and speciation is in agreement with proposals arising from work in other species, e.g. it has been suggested that speciation in *Drosophila* takes place particularly in relation to the appearance of novel sexual dimorphisms that are then subjected to a process of sexual selection (Kaneshiro 1980; Kaneshiro and Boake 1987; Carson 1997). Somewhat similar suggestions have been made for the colour changes that accompany the rapid speciation of cichlid fish that has taken place in the lakes of East Africa (Dominey 1984; McKaye 1991), and in relation to changes between species in song morphology and plumage in birds (Price 1998). In each case the focus of the speciation process is a sexual dimorphism (a difference in some physical or behavioural characteristic between the sexes).

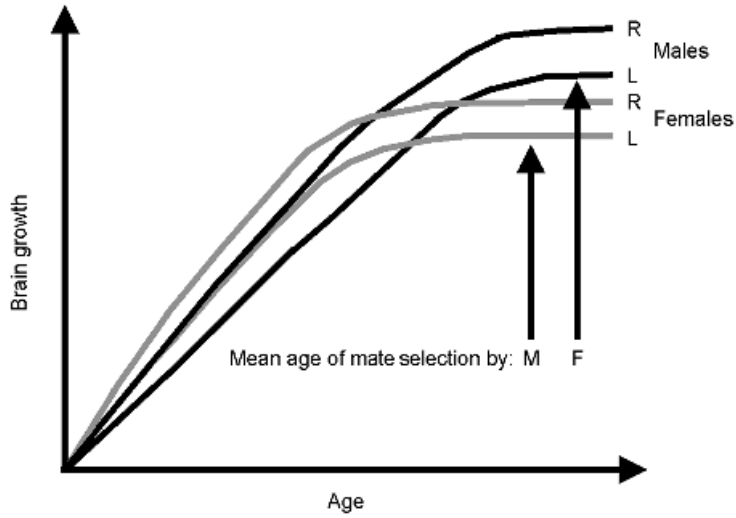


Figure 4.6. Hypothetical trajectories of growth of the cerebral hemispheres in man under the influence of an asymmetry determinant (Annett's 'right shift factor' or the cerebral dominance gene, located in homologous form on the X and the Y chromosome) acting early in development. Genetic (or epigenetic, i.e. gene expression rather than gene sequence) variation is associated with different trajectories of relative growth of the left (L) and right (R) hemispheres, the degree of asymmetry being determined by variation on the X and the Y chromosomes. Note that the asymmetry of the human brain is, in reality, more complex than a simple right–left difference; it is better represented as a 'torque' (a bias or twist; see Figure 4.3) across the anteroposterior axis from right frontal to left occipital. Mean age at mate choice differs between the sexes; females generally choose mates who are older than themselves and males generally choose females who are younger than themselves (see Crow 1993b)

Thus, the evolution of language in *Homo sapiens* (Crow 1996, 1998a,b) may be a specific case of a more general phenomenon. In *Homo sapiens* it is proposed (Crow 1993b) that the primary change (the change that led to language) occurred in a gene that influenced the relative development of the two hemispheres, and that this had an effect on the rate of brain growth. It is established that the rate of brain growth in man is different in the two sexes (e.g. Figure 4.5). Whether this sex difference is related to differences in the sequence of the gene on the X and the Y (such as are known to be present in the case of protocadherinXY), or to differences in residual inactivation on the X, remains to be investigated.

If the relative development of the two hemispheres is regulated by an X–Y homologous gene, this implies that the trajectory of brain growth is susceptible to differential modification in the two sexes. The target of selection is presumably related to the plateau of brain growth (Figure 4.6), the stage at which maturation

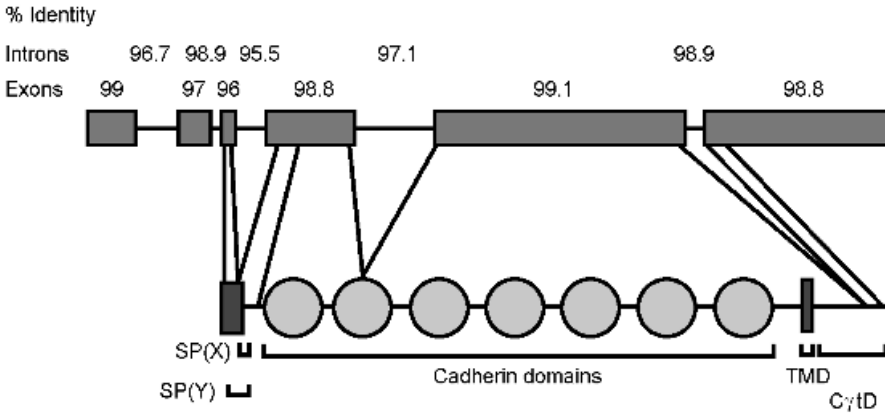


Figure 4.7. The structure of protocadherinXY. In the DNA structure (upper part of diagram) there are six exons (thick bars representing the part of the molecule that is transcribed into RNA and then translated into protein) separated by five introns (thin horizontal lines), each with 96–99.1% homology between X and Y copies. In protein structure (lower part of diagram) there are seven extracellular cadherin domains (filled circles, representing the part of the molecule on the cell surface that interacts with a similar molecule on the surface of another cell), a signal peptide (SP; concerned with the delivery of the molecule to its proper location on the cell membrane) that differs between the X and the Y copies, a transmembrane domain (TMD) that crosses the cell membrane, and a cytoplasmic domain (CytD; the mediator of effects within the cell) that again differs in termination sequence between X and Y copies. Each of the differences between X and Y forms of the protein has the potential to explain a difference between the sexes. Adapted from Blanco *et al.* (2000)

of the brain is reached. The point of selection as a mate of males by females may be assumed to be later than that of females by males, as reflected by the consistency across populations of the 2–3 year difference in age at marriage (Crow 1993b) — males are older than females at the time of marriage.

The preference of females for older males is consistent with the assumption that the peak of lateralisation and verbal ability occurs later in males than in females. If the target of selection is verbal ability, it appears that females select mates whose linguistic prowess matches their own, and that they are thereby determining the latest point at which optimal lateralisation may occur. With the introduction of the key dimension of lateralisation, strong selection of the linguistically able over the less able has the potential over time to bring about big changes in the point of plateau in brain growth.

Here one can see there are three interacting variables — degree of lateralisation, age of being selected as a mate, and the rate (perhaps determining the plateau) of brain growth. For each of these variables, the difference between the sexes may reflect differences in the sequence or

expression of a gene on the X and the Y chromosome. At what stage in hominid evolution were these sex differences introduced? It can be envisaged that the primary translocation from the X to the Y (that occurred 2–3 million years ago; Sargent *et al*. 1996), which doubled the gene dosage, prolonged brain maturation in males and initiated a sequence of changes, including one affecting the inactivation status of the genes in this region on the inactive X chromosome in females. There is some evidence for asymmetry in *Homo erectus* (Steele 1998) and this change might be relevant. One can speculate that the paracentric inversion that came later (but cannot be precisely dated; Schwartz *et al*. 1998) was more relevant to the change that brought about modern *Homo sapiens*.

Thus, a sequence of hypotheses can be formulated:

1. Lateralisation is the critical change that defines the human brain and has conferred upon it the capacity for language (Annett 1985; Corballis 1991).
2. This change occurred in the course of hominid evolution (Marchant and McGrew 1996; Buxhoeveden and Casanova 2000).
3. Given the neuropsychological findings in sex chromosomal aneuploidies outlined above, that this was the consequence of a change on the sex chromosomes.
4. The evolutionary history of these chromosomes points to the Xq21.3 to Y translocation as the key event.

Whether or not lateralisation followed the original translocation or occurred later (e.g. at the time of the paracentric inversion) the presence of the gene on the Y chromosome introduced the possibility of differential modification in the two sexes. The important point is that a change that creates a new region of homology between the X and the Y chromosomes establishes a new sexual dimorphism and sets up a situation in which sexual selection can act first on males and then on females to modify the relevant characteristic. This, it is proposed, is what happened following the Xq21.3/Yp translocation and its subsequent modifications, to generate hemispheric differentiation and the capacity for language. The fact that there are language anomalies as well as hemispheric deviations in the sex chromosome aneuploidies (Money 1993) reinforces the general case that language is related to an X–Y homologous gene.

Thus, the evolution of language in *Homo sapiens* may be an exemplar of a general rule that links sex linkage, sexual selection and speciation. According to this rule, a primary change in the Y copy of an X–Y homologous gene: (a) generates a new sexual dimorphism; and (b) is subject to female choice. The X–Y difference can then become the target of runaway (Fisherian) sexual selection.

PROTOCADHERINXY

Of the 30,000 estimated genes in the genome, any gene within the 3.5 Mb region (approximately 0.2% of the genome) in the Xq21.3 block that translocated to the Y chromosome after the separation of the hominid lineage would be of interest in relation to distinctively human characteristics, because the expression of such a gene will have changed relative to its expression in the great apes. With the case outlined above for an asymmetry determinant in the X–Y homologous class, this region acquires increased interest in relation to the evolution of the cerebral cortex. Given the paucity of genes on the Y chromosome, few functional genes would be predicted to exist, and there is no *a priori* reason to expect any one gene within this region to be expressed in the brain. However, in the course of sequencing the region, one gene has been found to be expressed in the brain *and* to have characteristics relevant to nervous system growth and development. This gene (protocadherinXY; see Figure 4.7) belongs to a class that codes for proteins that are expressed on the surface of subsets of neurones to act as axon guidance molecules (Blanco *et al.* 2000). Protocadherins are a subfamily of the cadherins, molecules that are expressed in a regionally specific manner in the central nervous system to identify specific brain nuclei, fibre tracts and layers within structures such as the cerebral cortex. They contribute to morphogenesis and tract formation, in part in a species-specific manner (see e.g. Redies 2000; Yagi and Takeichi 2000).

It is clear from its location in the genome that the control and expression of protocadherinXY has changed in the course of human evolution, and this must have had consequences for the structure of the human brain. Given the background of evidence for an X–Y homologous determinant of cerebral asymmetry, it is plausible, although it remains to be demonstrated by studies of variations in gene sequence and expression, that these changes were critical in the evolution of language. On the basis of its location, inferred function and evolutionary history, protocadherinXY is a candidate for a central role in the evolution of language.

CONCLUSIONS

1. Language is an embarrassment for gradualist evolutionary theory because, according to some authors, it requires a saltation, i.e. a discontinuous ‘speciation event’.
2. Cerebral asymmetry is present in *Homo sapiens* but absent in the chimpanzee.

3. There is a case that a gene for asymmetry is present in homologous form on both the X and the Y chromosome.
4. A chromosomal change between the chimpanzee and *Homo sapiens* established a new region of homology between the X and the Y chromosome; it constitutes a saltational change of a type that would be expected to be subject to sexual selection.
5. Within that region (Xq21.3/Yp), a protocadherin gene with forms that now differ on the X and Y has been identified.
6. This gene would be expected to influence brain development differently in the two sexes; it could be relevant to the mean faster brain growth and development of verbal ability in females.
7. The role of language in *Homo sapiens* elucidates a putative general role of sex chromosomal change and sexual selection in transitions to specific mate recognition systems.

REFERENCES

Affara, N., Bishop, C., Brown, W., Cooke, H. *et al*. (1996). Report of the second international workshop on Y chromosome mapping 1995. *Cytogenet. Cell Genet*. **73**, 33–76.

Annett, M. (1985) *Left, Right, Hand and Brain; the Right Shift Theory*. Erlbaum, London.

Bateson, W. (1894). *Materials for the Study of Variation, Treated with Especial Regard to Discontinuity*. Macmillan, New York.

Bear, D.M., Schiff, D., Saver, M., Greenberg, R. and Freeman, R. (1986). Quantitative analysis of cerebral asymmetry: fronto-occipital correlation, sexual dimorphism and association with handedness. *Arch. Neurol*. **43**, 598–603.

Bickerton, D. (1995). *Language and Human Behavior*. University of Washington, Seattle, WA.

Blanco, P., Sargent, C.A., Boucher, C., Mitchell, M. and Affara, N.A (2000). Conservation of PCDHXY in mammals: expression of human X/Y genes predominantly in brain. *Mammal. Genome* **11**, 906–914.

Butler, S. (1984). Sex differences in human cerebral function. *Prog. Brain Res*. **61**, 443–454.

Buxhoeveden, D. and Casanova, M. (2000). Comparative lateralization patterns in the language area of normal human, chimpanzee, and rhesus monkey brain. *Laterality* **5**, 315–330.

Carson, H.L. (1997). Sexual selection: driver of genetic change in Hawaiian Drosophila. *J. Heredity* **88**, 343–352.

de Condillac, E.B. (1746). *Essai sur L'Origine des Connaissances Humaines*. Trans. T. Nugent, as *Essay in the Origin of Human Knowledge*. AMS Press, New York, 1974.

Chomsky, N. (1972). *Language and Mind*. Harcourt, Brace, Jovanovich, San Diego, CA.

Corballis, M.C. (1991). *The Lop-sided Ape. Evolution of the Generative Mind*. Oxford University Press, New York.

Corballis, M.C., Lee, K., McManus, I.C. and Crow, T.J. (1996). Location of the handedness gene on the X and Y chromosomes. *Am. J. Med. Genet*. (*Neuropsychiat. Genet*.) **67**, 50–52.

Coyne, J.A. and Orr, H.A. (1998). The evolutionary genetics of speciation. *Phil. Trans. R. Soc. Lond. B* **353**, 287–305.

Crow, T.J. (1991). Protection from X inactivation. *Nature* **353**, 710–710.

Crow, T.J. (1993a). Origins of psychosis and the evolution of human language and communication. In S. Langer, J. Mendlewicz and J. Racagni (eds), *New Generation of Antipsychotic Drugs: Novel Mechanisms of Action*. Karger, Basel, pp. 39–61.

Crow, T.J. (1993b). Sexual selection, Machiavellian intelligence and the origins of psychosis. *Lancet* **342**, 594–598.

Crow, T.J. (1996). Language and psychosis: common evolutionary origins. *Endeavour* **20**, 105–109.

Crow, T.J. (1998a). Sexual selection, timing and the descent of man: a theory of the genetic origins of language. *Curr. Psychol. Cogn*. **17**, 1079–1114.

Crow, T.J. (1998b). Why cerebral asymmetry is the key to the origin of *Homo sapiens*: how to find the gene or eliminate the theory. *Curr. Psychol. Cogn*. **17**, 1237–1277.

Crow, T.J. (2000). Did *Homo sapiens* speciate on the Y chromosome? *Psycoloquy* **11**(01) (http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?11.001).

Crow, T.J., Crow, L.R., Done, D.J. and Leask, S.J. (1998). Relative hand skill predicts academic ability: global deficits of the point of hemispheric indecision. *Neuropsychologia*, **36**, 1275–1282.

Darwin, C. (1859). *The Origin of Species by Means of Natural Selection or The Preservation of Favoured Races in the Struggle for Life*. John Murray, London.

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*. John Murray, London, reissued in facsimile, Princeton University Press, NJ, 1981.

Dobzhansky, T. (1937). *Genetics and the Origin of Species*. Columbia University Press, New York.

Dominey, W.J. (1984). Effects of sexual selection and life histories on speciation: species flocks in African cichlids and Hawaiian *Drosophila*. In A.A. Echelle and I. Kornfield (eds), *Evolution of Fish Species Flocks*. Orono Press, Orono, ME, pp. 231–249.

Eldredge, N. and Gould, S.J. (1972). Punctuated equilibria: an alternative to phyletic gradualism. In T.M. Schopf (ed.), *Models in Palaeobiology*. Freeman Cooper, San Francisco, CA, pp. 82–115.

Freije, D., Helms, C., Watson, M.S. and Donis-Keller, H. (1992). Identification of a second pseudoautosomal region near the Xq and Yq telomeres. *Science* **258**, 1784–1787.

Geschwind, N. and Levitsky, W. (1968). Left–right asymmetry in temporal speech region. *Science* **161**, 186–187.

Goldschmidt, R. (1940). *The Material Basis of Evolution*. Yale University Press, New Haven, CT.

Halpern, D.F. (1992). *Sex Differences in Cognitive Abilities*. Erlbaum, Hillside, NJ.

Howard, D.J. and Berlocher, S.H. (eds) (1998). *Endless Forms: Species and Speciation*. Oxford University Press, Oxford.

Jegalian, K. and Page, D.C. (1998). A proposed mechanism by which genes common to mammalian X and Y chromosomes evolve to become X inactivated. *Nature* **394**, 776–780.

Kaneshiro, K.Y. (1980). Sexual isolation, speciation and the direction of evolution. *Evolution* **34**, 437–444.

Kaneshiro, K.Y. and Boake, C.R.B. (1987). Sexual selection: issues raised by Hawaiian *Drosophila*. *Trends Ecol. Evol*. **2**, 207–211.

King, M. (1993). *Species Evolution: the Role of Chromosome Change*. Cambridge University Press, Cambridge.
Kretschmann, H.-J., Schleicher, A., Wingert, F., Zilles, K. and Loeblich, H.-J. (1979). Human brain growth in the nineteenth and twentieth century. *J. Neurol. Sci.* **40**, 169–188.
Lambson, B., Affara, N.A., Mitchell, M. and Ferguson-Smith, M.A. (1992). Evolution of DNA sequence homologies between the sex chromosomes in primate species. *Genomics* **14**, 1032–1040.
Laval, S.H., Dann, J., Butler, R.J., Loftus, J. *et al*. (1998). Evidence for linkage to psychosis and cerebral asymmetry (relative hand skill) on the X chromosome. *Am. J. Med. Genet*. (*Neuropsychiat. Genet*.) **81**, 420–427.
Maccoby, E.E. and Jacklin, C.N. (1975). *The Psychology of Sex Differences*. Oxford University Press, Oxford.
McKaye, K.R. (1991). Sexual selection and the evolution of the cichlid fishes of Lake Malawi, Africa. In M.H.A. Keenleyside (ed.), *Cichlid Fish: Behavior, Ecology and Evolution*. Chapman & Hall, London, pp. 241–257.
McKeever, W.F. (2000). A new family handedness sample with findings consistent with X-linked transmission. *Br. J. Psychol.* **91**, 21–39.
Magurran, A.E. and May, R.M. (eds) (1999). *Evolution of Biological Diversity*. Oxford University Press, Oxford.
Marchant, L.F. and McGrew, W.C. (1996). Laterality of limb function in wild chimpanzees of Gombe National Park: comprehensive study of spontaneous activities. *J. Hum. Evol.* **30**, 427–443.
Mayr, E. (1963). *Animal Species and Evolution*. Harvard University Press, Cambridge, MA.
Mellars, P. (1998). Neanderthals, modern humans and the archaeological evidence for language. *Mem. California Acad. Sci.* **24**, 89–115.
Money, J. (1993). Specific neurocognitional impairments associated with Turner (45,X) and Klinefelter (47,XXY) syndromes: a review. *Soc. Biol.* **40**, 147–151.
Moore, T. (1967). Language and intelligence: a longitudinal study of the first eight years. *Hum. Devel.* **10**, 88–106.
Mumm, S., Molini, B., Terrell, J., Srivastava, A. and Schlessinger, D. (1997). Evolutionary features of the 4 Mb Xq21.3 XY homology region revealed by a map at 60 kb resolution. *Genome Res.* **7**, 307–314.
Netley, C. (1998). Sex chromosome aneuploidy and cognitive development. *Curr. Psychol. Cogn.* **17**, 1190–1197.
Noble, W. and Davidson, I. (1996). *Human Evolution, Language and Mind*. Cambridge University Press, Cambridge.
Otte, D. and Endler, J.A. (eds) (1989). *Speciation and Its Consequences*. Sinauer, Sunderland, MA.
Penner, J.G. (2000). *Evolution Challenged by Language and Speech*. Minerva, London.
Perelle, I.B. and Ehrman, L. (1994). An international study of human handedness: the data. *Behav. Genet.* **24**, 217–227.
Premack, D. (1985). 'Gavagai' or the future history of the language controversy. *Cognition* **19**, 207–296.
Price, T. (1998). Sexual selection and natural selection in bird speciation. *Phil. Trans. R. Soc. Lond.* (*Biol.*) **353**, 251–260.
Provins, K.A., Milner, A.D. and Kerr, P. (1982). Asymmetry of manual preference and performance. *Percept. Motor Skills* **54**, 179–194.
Redies, C. (2000). Cadherins in the central nervous system. *Progr. Neurobiol.* **61**, 611–648.

Sapir, E. (1921). *Language: an Introduction to the Study of Speech*. Harcourt Brace Jovanovich, New York.

Sargent, C.A., Briggs, H., Chalmers, I.J., Lambson, B. *et al*. (1996). The sequence organization of Yp/proximal Xq homologous regions of the human sex chromosomes is highly conserved. *Genomics* **32**, 200–209.

Schwartz, A., Chan, D.C., Brown, L.G., Alagappan, R. *et al*. (1998). Reconstructing hominid evolution: X-homologous block, created by X–Y transposition, was disrupted by Yp inversion through LINE–LINE recombination. *Hum. Mol. Genet*. **7**, 1–11.

Shucard, D.W., Shucard, J.L. and Thomas, D.G. (1987). Sex differences in electrophysiological activity in infancy: possible implications for language development. In S.U. Philips, S. Steele and C. Tanz (eds), *Language, Gender and Sex in Comparative Perspective*. Cambridge University Press, Cambridge, pp. 489–502.

Steele, J. (1998). Cerebral asymmetry, cognitive laterality, and human evolution. *Curr. Psychol. Cogn*. **16**, 1202–1214.

Stringer, C. and McKie, R. (1996). *African Exodus: the Origins of Modern Humanity*. Jonathan Cape, London.

de Vries, H. (1905). *Species and Varieties, Their Origin by Mutation*. Open Court, Chicago, IL.

White, M.J.D. (1973). *Modes of Speciation*. W.H. Freeman, San Francisco, CA.

Yagi, T. and Takeichi, M. (2000). Cadherin superfamily genes: functions, genomic organization, and neurologic diversity. *Genes Dev*. **14**, 1169–1180.

PART II

Culture and the 'Social Brain'

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

5

Evolution of the Cultured Mind: Lessons from Wild Chimpanzees

WILLIAM C. McGREW
Department of Anthropology and Zoology, Miami University, Oxford, OH, USA

The gist of this chapter is simple: I seek to show that real culture needs collectivity, that collectivity needs social intelligence, especially mind-reading, and that mind-reading needs a big brain. I argue that chimpanzees show all of these traits, to a minimal but sufficient degree. Thus, the apes can help us to understand the evolutionary emergence of human culture, for which our unique brain is a necessary precursor. However, first some groundwork must be laid.

CHIMPOLOGY

Chimpanzees (*Pan troglodytes*) have been studied in captivity for almost 90 years in settings ranging from household to zoo to laboratory (Ladygina-Kohts 2002; Köhler 1927). Recently, these settings have been more naturalistic, of groups living in spacious enclosures or sanctuaries (de Waal 1998; Whiten 2001). Chimpanzees in nature have been studied in depth for 40 years, most notably in the pioneering and ongoing research of Jane Goodall at Gombe, in western Tanzania (Goodall 1986; Nishida 1990; Boesch and Boesch-Achermann 2000). While some of these studies have been compromised by artificial feeding or crop-raiding or other human interventions, other projects have none of these constraints.

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

The chimpanzee (and its cousin, the bonobo) is the closest living relative of *Homo sapiens*, sharing more than 99% of its genome with us. Remarkably, it is genetically closer to humans than to the other African great ape, the gorilla, despite physical appearances.

Wild chimpanzees have been studied, ecologically, ethologically and genetically, throughout their wide range in equatorial Africa, from Tanzania to Senegal. Six populations (Bossou, Budongo, Gombe, Kibale, Mahale and Taï) can be observed at close range, all day long, day after day. We can do ethnography, on individuals and over generations, comparing these wild apes at various levels (Whiten *et al*. 1999, 2001).

Table 5.1 shows the six levels at which wild chimpanzees are studied. Individual differences in behaviour are obvious and well-documented, e.g. at Gombe, Frodo is a successful hunter but a social misfit, by comparison with his half-brother, Freud. Differences across matrilineal family lines are barely known, but likely to be important. Gombe's matriarch, Fifi, has a recognisable maternal style, and her reproductive success far exceeds that of any other mother (Goodall 1986). Her daughters show the same style. At Mahale, neighbouring groups, K and M, show variants in their performance of the grooming hand-clasp (McGrew *et al*. 2001). Within East Africa, the Ugandan and Tanzanian populations differ in subsistence technology. The latter (Gombe, Mahale) show a rich range, especially in harvesting social insects for food, while the former (Budougo, Kibale) show almost none of this (McGrew 1992). Across regions of Africa, the different subspecies differ in extractive percussive technology. Only the far-western subspecies, *P. t. verus*, uses hammer and anvil to crack nuts; the central and eastern subspecies have the nuts and raw materials but show no cracking (McGrew *et al*. 1997). Finally, the two sibling species of *Pan* differ in foraging technology: chimpanzees show a rich array and bonobos none, although they inhabit similar forests (McGrew 1992; Whiten *et al*. 1999).

For no other taxa of large-brained mammal is such a range of data available: dolphin or whale, elephant, or other non-human primate. Chimpanzees are simply the best living species with which to model the progression of human evolution, hence they are the focus of this chapter.

Table 5.1. Levels of chimpanzee behavioural diversity, from specific to general

Individual	Frodo hunting success
Familial	F-family child-rearing
Group	K- vs. M-group grooming hand-clasp
Population	Uganda vs. Tanzanian subsistence technology
Sub-species	Far West African nut-cracking
Species	Bonobo lacks foraging technology

BIASES

Like everyone else, I have my biases, and these constrain my thinking. It seems best to make these explicit, rather than leave the reader to guess. First, my training is in ethology and ecology, not in anatomy and physiology. Thus, I can judge brains only by their behavioural products. Second, I am a field worker who observes organisms in the messy context of nature. I appreciate the elegance of experimentation but am sometimes sceptical of its artificiality. Third, my education was in both the social and the natural sciences, and I see no clear division between them, as the key is the scientific method, not the species of creature targeted. Fourth, I am a chimpanzee specialist, not always able to know enough about other primates, much less about other taxa. Finally, my viewpoint is Darwinian evolution. I see no other way to explain brain and mind.

NON-HUMAN CULTURE

If culture is uniquely human, then we have no need of apes to help us understand its origins. Or, if culture is widespread among vertebrates, then we might as well choose the guppy or chaffinch or black rat as our model. They are more accessible, affordable, and overall convenient than the chimpanzee. Both of these caveats must be addressed.

Many humanists, social scientists and even natural scientists believe that culture is unique to *Homo s. sapiens* among all living species (cf. McGrew 1998; e.g. Tomasello 1999). (There is another debate about cultural evolution in human ancestors, i.e. before the emergence of behaviourally modern humans at 100,000–200,000 years ago, or even after that, with regard to Neanderthals.) They point to symphonies, microchips and evolutionary theory as evidence of an obvious cultural Rubicon, ignoring the point that many traditional human societies would fail to qualify as cultural on these ethnocentric grounds. For material culture, it is hard to find qualitative differences between, for example, the subsistence technology of Tasmanian aborigines and Gombe chimpanzees (McGrew 1987). For non-material culture, the hurdle is methodological: How to know the content of an ape's mind? How to infer attitude, knowledge or meaning in another species? Whoever solves this problem will deserve a Nobel Prize, but in the meantime, we tend to make anthropocentric assumptions, ignoring the maxim that 'Absence of evidence is not evidence of absence' (verbal report is *no* solution: if apes could speak, would their words be any more truthful than those of our fellow human informants?).

Of course, human culture is unique, but so is human digestion. Why do we so readily assume phylogenetic continuity in the latter but deny it *a priori* in the former?

Table 5.2. Black rats (*Rattus rattus*) exploiting seeds of *Pinus halepensis* cones

Innovation	Recent exploiting of new resource
Dissemination	Technique passes from skilled to naïve
Standardisation	Two energy-efficient techniques: strip or shave
Durability	Offspring learn from mother's processing
Diffusion	Two nearby forests aged <50 years, each of 15 ha
Tradition	Well-established through multiple generations

From Terkel (1996) and Kroeber (1928).

Researchers who study species other than apes often go to the other extreme, of granting cultural status to all primates, or to mammals, vertebrates or even invertebrates. Their criteria for doing so differ. Bonner (1980) was willing to grant some sort of cultural status to slime-moulds on the grounds of their sociality. Others require the demonstration of social influence in behaviour, as in the case of the famous milk bottletop-opening blue tits of the British Isles (Whiten and Ham 1992; Whiten 2000). Most scientists grant cultural status on the basis of social learning, making it the necessary and sufficient condition to be met: on this criterion, the octopus may qualify, although it lacks a brain. Another standard that satisfies some students of culture is the demonstration of tradition. In these terms, if behavioural continuity across generations can be seen, then culture is shown.

Table 5.2 presents a case study of a rodent candidate for culture that meets all of these criteria and more. Terkel (1996) and his students studied wild black rats in Israeli pine forests and then in the laboratory. These opportunistic small mammals occupied the 'empty niche' of the squirrel in a plantation, becoming secondarily arboreal. They invented two techniques of harvesting pine kernels. The mother's technique passed to the offspring and then spread across forests, becoming well-established. Thus, these rats satisfied all six of the criteria implicit in Kroeber's (1928) provocative early work: innovation, dissemination, standardisation, durability, diffusion, tradition. How can these creatures be denied culture?

The rats have not been shown to be cultural because they lack society. That is, sociality is not enough; a fully cultural creature must have collective awareness. All that the rats show, however clever and adaptive, can be explained by one-to-one social learning of familial ways. This key distinction is addressed in the next section.

COLLECTIVITY

Consider a shoal of herring. It moves in perfect unison, diving and wheeling as one unit in precision. All members' movements must be coordinated, and their acts are likely to be beneficial in reducing predation. Each could choose to live

alone, merely by swimming a solitary course, but instead opts for safe and constant sociality. Surely, this is a natural collective?

Actually, it is not, at least in my terms. Nothing in the herring's behaviour suggests that it is more than a perceptually-reactive automaton, embedded in the geometry of a selfish herd (Hamilton 1971). The almost choreographic group movements may be no more than the sum of anonymous (and selfish) individuality acting in unison. Thus, the result is no more a collective than is the coordinated positioning of a trillion grains of sand in a wind-sculpted dune. (I acknowledge that, in my ignorance of herring mentality, this picture may be incomplete. I stand ready to be corrected if contradictory data come forward.)

What is missing in the herring is any indication that an individual is aware of being a member of the group. Instead of relationships based on individual recognition, there is apparent interchangeability of conspecifics, which might as well be strangers. There is no sign of collective consciousness, even in the most minimal sense of conceptual vs. perceptual responses to stimuli. Put another way, there is no sign of self-awareness, much less of social identity. Obligatory sociality here means no more than moment-to-moment spatiotemporal synchrony that requires only the ability to distinguish between own and other species. Norms (in a statistical sense) there may be, but no sign of standards (agreed-upon thresholds or criteria).

How to translate such ideas into empirically testable hypotheses? How to measure collectivity or identity? There follow some heuristic examples known or thought to exist in wild chimpanzees.

At Mahale, an alpha male sets the tone for distribution of prized resources, most notably for the sharing of meat after a kill of a red colobus monkey (Nishida 1992). This is a *role*, assumed upon the achievement of the top position and lost when another usurps it. There is nothing (known to be) intrinsic to any individual about how to distribute meat, but it affects the whole community's consumption, and thus its collective response to a hunt. Recent findings show that this collective milieu changes when the alpha's role passes from one male to another (Marchant 2002). The new alpha's style (personality?) reconfigures the process, and thus the group's structured interaction. Thus, at Mahale, Ntologi's Machiavellian differential distribution of meat to allies, rivals and neutrals gave way to Nsaba's (crude?) two-male oligarchy that denied meat to all other males.

At Tai, Boesch and Boesch-Achermann (2000) and Boesch (2002) described *division of labour* in the social hunting of red colobus monkeys. They told how, in any given hunt, an individual might be a Driver ('herding' the prey toward other hunters), a Blocker (cutting off the prey's potential escape routes) or an Ambusher (waiting to subdue the prey when driven close enough). Others have offered alternative explanations, of each hunter positioning himself selfishly, filling holes, so to speak, in order to maximise being the one who achieves first possession of the kill (Stanford 1998). In any event, such a hunt is a collective,

interdependent set of acts; there is no point in having an Ambusher without a Driver, or vice versa. What needs to be shown is whether the division of labour is one of roles (turn-taking?) or intrinsic determinants (agility for ambushing, patience for blocking, etc.) or occurs by chance.

At Gombe, collectivity is expressed most starkly in *xenophobia* that leads to fatality (Goodall *et al.* 1979; Wrangham and Peterson 1996). Parties of males patrol the boundaries of the community's territory and attack strangers found in these border zones. Sometimes they travel into the neighbour's territory, on raids. If they encounter a female, she may be beaten or kidnapped and her offspring slain; a male may be killed outright or fatally wounded. This is the ultimate in solidarity, of 'us' vs. 'them'. However, there is more to this than intergroup aggression. In the best-documented case, one community, Kasakela, annihilated another, Kahama, although the latter had formerly been their close associates. Thus, the seceding Kahama males transformed their collectivity, and went from insiders to outsiders, with lethal consequences. Goliath, once the alpha male of the combined community, was brutally dispatched by his former companions (Goodall 1986). Thus, assumed or shed identity by the Kahama males was extrinsically, not intrinsically, determined. In effect, they chose to redefine themselves and paid the price.

Also at Mahale, the chimpanzees show a social *convention*, i.e. a standardised and arbitrary enhancement of a species-typical activity. This is the grooming hand-clasp (GHC), which is grafted on to normal social grooming (McGrew and Tutin 1978). Chimpanzees everywhere engage in unilateral, mutual and reciprocal social grooming — this is a chimpanzee universal (Whiten *et al.* 1999). At Mahale they also regularly show a GHC which consists of two individuals sitting facing one another in mirror-image configuration, each with one arm fully extended, hands clasped overhead, while the other hand grooms the revealed armpit. The resulting 'A-frame' posture is striking. It serves no direct function, i.e. it does not increase the hygienic efficiency of grooming, and so has been labelled as a 'social custom'.

Later comparison of neighbouring groups within the Mahale population revealed more complexity (McGrew *et al.* 2001); whereas K-group performs the classic, symmetrical GHC, M-group shows asymmetry: one of the participant's hands provides the support (does the work), while the other's hand is supported (gets a free ride). The difference is not trivial in terms of effort (energy); try holding your laptop overhead in one hand, arm fully extended, for more than a couple of minutes!

So, why do the GHC at all? There is a 100% correlation between who supports vs. who is supported, and their respective social ranks. Subordinates hold up dominants: when alpha grooms with beta, beta supports; when beta grooms with gamma, beta is supported. Thus, the outstretched arms provide an obvious and unambiguous acknowledgement of rank within the dyad and signal the same to others in the group. Hence, GHC appears to function for

chimpanzees as does a military salute in humans. Such conventions imply collective consciousness (for a similar line of argument applied to another aspect of social grooming, the social scratch, see Nakamura *et al.* 2000).

These are four types of behaviour from three populations of chimpanzees, all of which may imply a dimension of sociality in which the common factor is collectivity. Similar phenomena are known from other taxa, such as cooperatively breeding birds or large social carnivores, but these species seem to lack such individual differences and multi-dimensionality (McGrew 2002).

I argue that such collectivity is qualitatively different from what has been labelled as 'culture' in other non-human species, whether these be song-learning passerines, sweet-potato-washing macaques, or pinecone-processing rats. If this is so, then it demands explanation in terms of cognitive capacity.

THE MENTALITY OF CULTURE

Non-social learning, by whatever mechanism, requires no intentionality and is correspondingly widespread in organisms. Social learning, in the broad sense of Whiten and Ham (1992) including social influences, requires first-order intentionality ('I know') (Dennett 1981). This is so because of the dynamic contingency of social interaction, in which participants modify or modulate their reactions in relation to the actions of the other (actually, this may also occur in some predator–prey interactions and so is not strictly social, i.e. conspecific). This occurs whether the interactant is stranger or familiar, friend or foe, kin or non-kin. To learn from another means, minimally, being able to accommodate one's actions *vis-à-vis* another's; this is crucially different from social encounters in which no information is passed that alters the recipient's knowledge-base. A chimpanzee who takes nuts cracked by another is only a thief; a chimpanzee who steals another's nut-cracking hammer is a better-informed thief; but a chimpanzee who purloins another's nut-cracking technique is an intentional thief.

By this line of argument, collectivity, as part of fully-blown culture, requires second-order intentionality ('I know that you know'), This is empathy, also known as mind-reading or 'theory of mind' (Whiten 1996). Schiefenhövel (this volume) has termed this combination of perspective taking and empathy as *enphronesis*. It means being able to put yourself in another's position, mentally. To assume another's role may require empathy, as may (voluntary) participation in division of labour, transformational xenophobia and collaboration in a conventionalised ritual. But each of these is debatable, and may be explained more parsimoniously.

Could any aspect of chimpanzee social life unequivocally indicate second-order intentionality?

TEACHING

By definition, teaching entails second-order intentionality in a slightly modified form: 'I know that you don't know' (Caro and Hauser 1992). Unlike all other forms of information transmission, it is a necessary condition of teaching that the tutor be aware of the state of (lack of) knowledge of the pupil. This is what distinguishes teaching from display and from training. In display, the performer seeks to influence the other's actions. In training, the trainer seeks to modify the other's actions. In teaching, the teacher seeks to inform the other's actions. Put another way, trainers enact protocols of reward or punishment that need not take into account the state of knowledge of the trainee. Teachers, on the other hand, assess the initial knowledge-state of the pupil, then tailor their acts accordingly, modifying these as progress is (or is not) made. Teaching has to be empathetic.

Teaching in wild chimpanzees remains unconfirmed. Boesch (1991) showed training in nut-cracking at Tai, but his data for teaching were anecdotal. Reports from other sites are similarly sparse (Goodall 1986).

There is no need to explain negative results, but to do so can be informative, at least with regard to hypothesis-posing. Like humans, apes should turn to teaching only as a last resort, when all other forms of information transfer are inadequate. Why? Because teaching incurs costs as well as benefits, whether in time, energy or risk, just like any other social interaction. In all other forms of social learning, the onus is on the learner to incur these costs, while the knowledgeable individual goes about its normal business.

Table 5.3 shows the four logical results of pedagogy, following Trivers (1985). I know of no quantitative test of the relative frequency of these four outcomes, but they can be surmised. Most likely is altruism, in which the teacher invests in the pupil at no little cost to the teacher's activity budget or emotional reserves (have you ever tried to teach someone to drive a car?). Such altruism in apes seems tenable only as parental or kin investment.

The other three alternatives seem less likely, but for different reasons: spiteful teaching might entail taking time, effort or risk to teach someone poorly or even wrongly. Thus, both parties pay a net cost. This might occur to forestall a future competitor, as a sort of mal-apprenticeship.

Table 5.3. Sociobiology of pedagogy (after Trivers 1985): four logical outcomes to a teaching interaction

		Pupil	
		Net benefit	Net cost
Teacher	Net benefit	Cooperation	Competition
	Net cost	Altruism	Spite

Competitive teaching might entail teaching kin but not non-kin, or close-knit kin but not distant kin, or reciprocators but not cheats. Or, competitive teaching might entail extracting 'payment' from a pupil, when the cost to the latter of remaining ignorant is even higher.

Finally, cooperative teaching is expected, if an informed companion would make a more productive contribution to a cooperative effort, e.g. hunting. (Note, however, that this entails investment in advance of the payoff, risking that the well-tutored recipient will not defect. Simultaneous, cooperative teaching is hard to envision.)

Thus, one might expect to see teaching in situations where doing something precisely right is important, such as technological survival skills, as in nut-cracking, or social skills that affect reproductive success, as with dominance signals. Since these are uncommon in nature, experiments in captivity may be called for. I know of no experimental studies in which an individual chimpanzee has been rewarded for teaching (as opposed to informing) another.

In sum, there are multiple strands of evidence of mind-reading in wild chimpanzees, as inferred from observational data. The way is clear for experimental studies in captivity to probe these capacities further.

BRAIN AND BEHAVIOUR

Such complex software as full-blown culture requires co-evolved hardware. Gross comparisons of brains across primates, and especially within the Hominoidea, suggest major differences: on average the brain of *Homo sapiens* is 3–4 times the volume of the brain of *Pan troglodytes*.

Functional comparisons, especially using modern imaging technology, yield surprising similarities, especially with regard to cerebral asymmetries associated with complex social cognition. Gannon *et al.* (1998) showed that the planum temporale, which is a key component of Wernicke's receptive language area, is larger on the left hemisphere of chimpanzees, just as it is in humans. Fernandez-Carriba *et al.* (2002) showed that chimpanzees show a right-hemisphere specialisation for the innervation of facial expression of emotions, just as in humans. Both of these examples force reappraisal of mental abilities previously thought to be unique to hominids.

How such a brain came to be has been much debated, especially since Humphrey's (1976) seminal paper hypothesised that social, not subsistence, selection pressures were the driving force in the evolution of intelligence. Byrne and Whiten (1988) focused on tactical deception as the crucial element, leading to what they called 'Machiavellian intelligence'. Dunbar (1998) summarised what has come to be known as the 'social brain' hypothesis, and tested empirically correlations between cerebral characteristics and social variables.

Despite the relative crudeness of the measures, e.g. group size as a proxy for social complexity, lawful relationships emerged from these analyses.

This is not to say that the alternative hypothesis of subsistence as the key set of selection pressures, whether in dietary choice or food processing, has been abandoned. Byrne (1995) reminded us that the 'food-for-thought' hypothesis can explain aspects of intelligence that are common to great apes who differ markedly in their social structure. Chimpanzees and bonobos, with their promiscuous, fission–fusion community structure, differ notably from gorillas, with their tightly bound harem groups, yet are found to be of similar intelligence by all known systematic comparative studies. A similar argument applies to the orang utan, whose essentially solitary lifestyle could hardly be more different from its African cousins' varieties of sociality. What all living hominoids have in common is complex foraging strategies, from mental maps of resources to extractive manipulatory techniques. This is congruent with subsistence as the adaptive force driving intelligence.

However, another aspect of subsistence, the use of elementary technology, is murky to the point of bewildering. Chimpanzees and orang utans in nature employ tools to make a living; bonobos and gorillas do not. In captivity, all four species will use tools with equal skill if it is made worth their while to do so. This is confusing (McGrew 1989).

So, what of the 'subsistence' brain? Apart from the crudest comparisons on grounds of dietary choice (e.g. frugivory vs. folivory), little has been done to match the comparable analyses of sociality across the primate order. There is not enough variation in elementary technology, as it is largely absent in non-hominoid primates, and not enough known about, for example, mental maps in other taxa to allow correlational testing.

Surprisingly, even the brain data are incomplete, with two of the four species of great apes, orang utan and bonobo, absent from most comparative analysis (e.g. Dunbar 1998).

In sum, the available data on cerebral functioning do not clearly support either the 'social' or 'subsistence' brain hypothesis. The more we learn about the neural substrate for social cognition, however, the more impressively similar is the brain–behaviour connection in chimpanzees and humans.

CONCLUSIONS

In this chapter, the case has been made that we now know enough about chimpanzees in nature to tackle questions reflecting complex phenomena such as culture and cognition. A constrained, stricter definition of culture, based on the feature of collectivity, takes us beyond more general processes, such as social learning and tradition. The result is an apparent insight, that such fully developed culture requires equally derived cognition in the form of

mind-reading. In this area, as in so many others, chimpanzees are more like humans than they are like other primates, at least on the basis of currently available evidence. Thus, the key to culture is not just intelligence or adaptability but enphronesis, and this is apparently shared by humans and their nearest relations.

ACKNOWLEDGEMENTS

I thank the organisers for their invitation to take part in the Bochum conference, L. F. Marchant for critical comments on the manuscript and D. L. Deaton for manuscript preparation. Special acknowledgement goes to C. E. G. Tutin for her help at the earliest state of my awareness of nonhuman primate culture, more than 25 years ago.

REFERENCES

Boesch, C. (1991). Teaching among wild chimpanzees. *Anim. Behav.* **41**, 530–532.
Boesch, C. (2002). Cooperative hunting voles among Taï chimpanzees. *Hum. Nat.* **13**, 27–46.
Boesch, C. and Boesch-Achermann, H. (2000). *The Chimpanzees of the Taï Forest*. Oxford University Press, Oxford.
Bonner, J.T. (1980). *The Evolution of Culture in Animals*. Princeton University Press, Princeton.
Byrne, R.W. (1995). *The Thinking Ape*. Oxford University Press, Oxford.
Byrne, R.W. and Whiten, A. (eds) (1988). *Machiavellian Intelligence*. Oxford University Press, Oxford.
Caro, T.M. and Hauser, M.D. (1992). Is there teaching in non-human animals? *Qu. Rev. Biol.* **67**, 151–174.
Dennett, D. (1981). *Brainstorms*. MIT Press, Cambridge, MA.
Dunbar, R.I.M. (1998). The social brain hypothesis. *Evol. Anthropol.* **6**, 178–190.
Fernandez-Carriba, S., Loeches, A., Morcillo, A. and Hopkins, W.D. (2002). Functional asymmetry of emotions in primates: new findings in chimpanzees. *Brain Res. Bull.* **57**, 561–564.
Gannon, P.J., Holloway, R.L., Broadfield, D.C. and Braun, A.R. (1998). Asymmetry of chimpanzee planum temporale: human-like pattern of Wernicke's brain language area homology. *Science* **279**, 220–227.
Goodall, J. (1986). *The Chimpanzees of Gombe*. Harvard University Press, Cambridge, MA.
Goodall, J. *et al.* (1979). Intercommunity interactions in the chimpanzee population of the Gombe National Park. In D.A. Hamburg and E.R. McCown (eds), *The Great Apes*. Benjamin/Cummings, Menlo Park, CA, pp. 13–53.
Hamilton, W. (1971). Geometry for the selfish herd. *J. Theoret. Biol.* **31**, 295–311.
Humphrey, N.K. (1976). The social function of intellect. In P.P.G. Bateson and R.A. Hinde (eds), *Growing Points in Ethology*. Cambridge University Press, Cambridge, pp. 303–321.
Köhler, W. (1927). *The Mentality of Apes*. Kegan Paul, Trench, Trubner, London.
Kroeber, A.L. (1928). Sub-human cultural beginnings. *Qu. Rev. Biol.* **3**, 325–342.

Ladygina-Kohts, N. (2002). *Infant Chimpanzee, Human Child*. Oxford University Press, New York.
McGrew, W.C. (1987). Tools to get food: the subsistants of Tasmanian aborigines and Tanzanian chimpanzees compared. *J. Anthropol. Res.* **43**, 247–258.
McGrew, W.C. (1989). Why is ape tool-use so confusing? In V. Standen and R.A. Foley (eds), *Comparative Socioecology*. Blackwell, Oxford, pp. 457–472.
McGrew, W.C. (1992). *Chimpanzee Material Culture*. Cambridge University Press, Cambridge.
McGrew, W.C. (1998). Culture in non-human primates? *Ann. Rev. Anthropol.* **27**, 301–328.
McGrew, W.C. (2002). Ten dispatches from the chimpanzee culture wars. In F.B.M. de Waal and P.L. Tyack (eds), *Animal Social Complexity*. Harvard University Press, Cambridge, MA.
McGrew, W.C., Ham, R.M., White, L.T.J., Tutin, C.E.G. and Fernandez, M. (1997). Why don't chimpanzees in Gabon crack nuts? *Int. J. Primatol.* **18**, 353–374.
McGrew, W.C., Marchant, L.F., Scott, S.E. and Tutin, C.E.G. (2001). Inter-group differences in a social custom of wild chimpanzees: the grooming hand-clasp of the Mahale Mountains, Tanzania. *Curr. Anthropol.* **42**, 148–153.
McGrew, W.C. and Tutin, C.E.G. (1978). Evidence for a social custom in wild chimpanzees? *Man* **13**, 234–251.
Marchant, L.F. (2002). *Chimpanzee Carnivory: Why Eat Meat?* Miami University, Oxford, OH (VHS Video).
Nakamura, M., McGrew, W.C., Marchant, L.F. and Nishida, T. (2000). Social scratch: another custom in wild chimpanzees? *Primates* **41**, 237–248.
Nishida, T. (ed.) (1990). *The Chimpanzees of the Mahale Mountains*. University of Tokyo Press, Tokyo.
Nishida, T. (1992) Meat-sharing as a coalition strategy by an alpha male chimpanzee? In T. Nishida, W.C. McGrew, P. Marler, M. Pickford and F.B.M. de Waal (eds), *Topics in Primatology, vol. 1: Human Origins*. University of Tokyo Press, Tokyo, pp. 159–174.
Stanford, C.B. (1998). *Chimpanzee and Red Colobus*. Harvard University Press, Cambridge, MA.
Terkel, J. (1996). Cultural transmission of feeding behavior in the black rat (*Rattus rattus*). In C.M. Heyes and B.G. Galef (eds), *Social Learning in Animals*. Academic Press, London, pp. 17–47.
Tomasello, M. (1999). The human adaptation for culture. *Ann. Rev. Anthropol.* **28**, 509–529.
Trivers, R.L. (1985). *Social Evolution*. Benjamin/Cummings, Menlo Park, CA.
de Waal, F.B.M. (1998). *Chimpanzee Politics*, revised edn. Johns Hopkins University Press, Baltimore, MD.
Whiten, A. (1996). Imitation, pretence, and mindreading: secondary representation in comparative psychology and developmental psychology? In A.E. Russon, K.A. Bard and S.T. Parker (eds), *Reaching into Thought*. Cambridge University Press, Cambridge, pp. 300–324.
Whiten, A. (2000). Primate culture and social learning. *Cogn. Sci.* **24**, 477–508.
Whiten, A. (2001). The evolutionary roots of culture. *Br. Acad. Rev.* **July–Dec**, 57–60.
Whiten, A., Goodall, J., McGrew, W.C., Nishida, T. *et al.* (1999). Cultures in chimpanzees. *Nature* **399**, 682–685.
Whiten, A., Goodall, J., McGrew, W.C., Nishida, T. *et al.* (2001). Charting cultural variation in chimpanzees. *Behaviour* **138**, 1481–1516.
Whiten, A. and Ham, R. (1992). On the nature and evolution of imitation in the animal kingdom: reappraisal of a century of research. *Adv. Stud. Behav.* **21**, 239–283.
Wrangham, R.W. and Peterson, D. (1996). *Demonic Males*. Houghton Mifflin, Boston, MA.

6

Ninye Kanye: the Human Mind. Traditional Papuan Societies as Models to Understand Evolution towards the Social Brain

WULF SCHIEFENHÖVEL
Human Ethology Group, Max-Planck-Society, Andechs, Germany

ENPHRONESIS, UNDERSTANDING THE MIND OF ANOTHER PERSON

'Theory of mind' (ToM; Premack and Woodruff, 1978) is the most common, yet, in my view, not very well-chosen term to express a brain's capacity to put itself into another brain, understand the mental processes at work there and deduce, from the knowledge thus gained, what own strategies would be best to answer or prevent the ones detected in this way. 'Mentalising' (mentalism, mentalist, are terms in philosophy and psychology referring to a very different concept) and 'mind reading' are other *termini technici* for this capacity, which I would like to name 'enphronesis', in accordance with the term 'empathy'. Enphronesis is based on the Greek word 'phren' (φρην) for 'diaphragma', which the Greeks believed was the seat of the mind; 'phren' therefore is 'mind', 'imagination', 'power of the mind' (the opposite is 'thymos' = 'heart', 'soul'...); 'phroneo', the verb, translates as: 'perceive mentally', 'discern mentally', 'think', 'understand', 'know', and the noun 'phronesis' (another form is 'phronema') means 'thinking', 'mind', 'intelligence', 'wisdom', 'insight', i.e. the capacity to understand the mind, the thoughts of another person.

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

When in phylogeny, did this remarkable ability start? Simple forms of enphronesis might perhaps be found in fishes or reptiles, and are probably not easily defined sufficiently and distinguished from other, more simpler forms of reacting to another individual's state of mind. From the work of William McGrew and Richard Byrne (this volume), Frans de Waal (1982) and others, it has become very clear that primates other than ourselves have a capacity for enphronesis and are able to 'read', and also to deceive, conspecifics. It is most likely, as with all other capacities, that there are precursors in phylogeny. We probably only have to look closely enough.

My contribution centres around observations I have made, since my first fieldwork in 1965/66 and many consecutive stays since then, in mainland and island New Guinea. Particularly, I will base my analysis on data collected in 22 months of fieldwork (1974–1976, 1980) among the Eipo, an ethnic group of the mountains of Papua, living just north of the central cordillera of West-New Guinea, recently officially named Propinsi Papua, Province of Papua.

THE EIPO AND THEIR ENPHRONESIS-PRONE WORLD

At the beginning of our interdisciplinary project in their midst, the Eipo lived an almost uninfluenced neolithic life in the pronounced isolation of their steep mountain valley, the valley of the Eipomek, or Eipo River. They called themselves 'Eipo' or 'Eipodumanang', the ones at the banks of the Eipo. The Eipo and their neighbours in the east, south, west and north belong to the Mek group of languages and cultures (Schiefenhövel 1976, 1991) and represent one of the numerous groups of the mountain regions of Papua.

Only few persons from outside had ever reached the Eipo Valley prior to the onset of our research in 1974. They include Pierre Gaisseau and his small team of journalists and adventurers, in their remarkable expedition crossing west New Guinea from the southern to the northern coast (Saulnier 1960); and again in 1969 Pierre Gaisseau, this time in a team of Indonesian military personnel parachuting into the upper Eipomek Valley (Laporan 1969); plus a couple of missionaries of the Unevangelised Fields Mission (UFM), who staked out, among other areas of the Daerah Jayawijaya (the highland sector of western New Guinea), this region as their territory for mission work. Other parts of the Mek area, especially in the North, remained uncontacted until after our project had been established in the upper Eipomek Valley and members of our team walked to some villages not visited before by outsiders. The Eipo and their Mek neighbours were not a small, isolated group, but a sizeable population of several thousand people, using their stone-age tool kit for subsistence and other activities, and who followed their traditional, often religiously meaningful, customs in our presence, as they had realised that we

had not come to change their lives but to observe and document it — which they, once we had become accepted, agreed to with pride.

At first glance it may seem that a culture that is as simple as that of the Eipo may not be a good candidate for tracing the origins of enphronesis as there is, for example, no sophisticated technique of presenting the mind of others in books, pictures, films, etc. Yet enphronesis has been shaped in the environment of evolutionary adaptedness, and for this social and ecological environment, I believe the quasi-neolithic life of the Eipo to be a very good model. The Eipo can indeed be seen as 'modern models of the past' (Schiefenhövel 1999).

In the possible scenarios that may help to explain which selection forces have brought about the human brain, with all its remarkable capacities, the tool-making and tool-using hypothesis has lost ground to the 'social brain' hypothesis, formulated by Brothers (1990). Many researchers now assume that the complexity of social groups, more precisely the interactions of their members, have formed sufficient (additional?) selection pressure to produce enphronesis, as it was of great adaptive value to be able to predict the possible moves of one's similarly brainy interaction partner.

The Eipo society, even though rather limited in size (approximately 800 persons), can serve as a good example for this basic assumption, as all members, even children, have, in this classic face-to-face society, a highly developed knowledge and notion of the Other. Everybody knows the name, family and parts of the genealogy of everybody within the village (40–200 persons) and, to a high degree, also in the political alliance of villages (approximately 600 persons) which team together for various activities, the most important of these being warfare against the hereditary enemy in the In valley to the West. Even the people in those enemy villages are often known by name. Particularly, they know the inhabitants of the villages in the upper Hei valley, a demanding trip across a pass at 3700 m above sea level to the Southern side of the central chain, where traditional trading partners and often marriage partners live.

I estimate that an ordinary adult member of a given Eipo village personally knows at least 500 people; of many others, including her/his dead relatives or relatives in other valleys, names and parts of the life history will also be known. Some Eipo, i.e. those with an exceptional memory, will know many more individuals, again either personally or by hearsay — which need not be an imprecise mode of transmission at all; on the contrary, many of my informants were able to give me the details (counterchecked through interviews with others) of the lives of dead relatives or of people, dead or living, in other regions, where they themselves had never been.

It was the 'academics' of the Eipo society, characterised by a large thirst for knowledge and analytically working minds, who were our teachers, driven by a strong pedagogical ethos. They managed to instil insight into the minds of their curious but (especially in the beginning) language-handicapped white visitors. I

affectionately remember how some of the juveniles and men spent whole nights in our hut, clad in empty rice bags to make up for the missing fire, not giving up their intense and vivid explanations until there were signs that the ethnographer had understood at least the gist of whatever complex matter was the issue. The surprising side was that these informants were not only able to explain even the most complicated conceptual, mythical or ritual item of their culture, but that they were also able to think through our skulls and detect which way would be the best to get their story across to our brains. After all, we spoke very little of their highly sophisticated Papuan language at first (knowledge had to be acquired monolingually), so another problem for them was to find ways of circumventing our linguistic inadequacy — enphronesis at full swing. In those nights, on the bark floor of the hut at the edge of Munggona village and in the light of a kerosene lamp, the power of the human mind became materially tangible, became transformed from the words of an indigenous person into my mind, and put into letters on the pages of my notebook and little tiny paper index cards, which I had cut out neatly to fit into little cardboard boxes . . . it was the time before the laptop.

One result of Eipo enphronesis and our efforts to match their performance is the dictionary of the Eipo language written by Volker Heeschen and myself (1983) (see Table 6.1 for some extracts from this). It contains not only 6000 words and their meanings, often covering a wide semantic spectrum, but also proverbs, sayings and *ad hoc* expressions, thus allowing a glimpse of the rich mental culture of these quasi-stone-age people. The dictionary is, I believe, printed proof that it is indeed, quite contrary to the stance of postmodern anthropology, very possible to bring back 'truth' from ethnographic endeavours.

Although many of these lexems have a self-reflexive mode (e.g. the words for being happy, sad, startled, unconscious, etc.), a number of them contain enphronesis aspects, e.g. the words for 'understanding', 'intelligence', 'to lie/deceive', 'to compare/give an example', 'to feel derision/malicious pleasure', 'to falsely accuse somebody', 'to be forgiving of injustice', and 'purposeful behaviour/plan'. It would be an interesting exercise to screen the languages of the world, especially those of small, hitherto isolated societies, for terms expressing the metacognitive vocabulary. To my knowledge, this has not yet been done.

EXTENDED FAMILIES AND COMPLICATED GENEALOGIES

Owing to the clear patrilineal descent rule, every Eipo person belongs to one of several clans passed on to him/her by the father. This patrilineage is the most precisely known axis of descent, important for land and other rights and for taboos and vital for survival because, even during visits to faraway villages, where brothers, fathers, cousins, etc. are absent and one is potentially in great

Table 6.1 Some entries, in alphabetical order, for the lexem *kanye* (= mind) in the Eipo language

kanye, *kanya*	Shadow, echo, image, thought, understanding, idea, mind, spirit, spirit of the deceased, of animals or features in nature
kanye barib-	To be in a depressed mood, to mourn, to think longingly; to find a place to live
kanye betinye	Two souls, i.e. to be in doubt
kanye bikina	Understanding, knowledge (*bik*, to know)
kanye bindohmanil	The soul/spirit leaves me, i.e. I became afraid
kanye bisik	Fossa supraclavicularis (the groove above the collar bone, where the soul is thought to leave the body during unconsciousness or after death)
kanye bobmal	He/she shows intelligence
kanye bobatek-	To be confused
kanye bobuk	He/she becomes intelligent, has understanding (e.g. children when they begin to speak)
kanye gum bobuk	He/she has lost his/her mind, has no brains
kanye dalolamak	They become frightened (e.g. when suddenly meeting someone)
kanye deib-	(Lit. to put the mind down) to find a place to live
kanye deibmanil	The mind is born to me; I am happy
kanye dib'lamak	The mind/spirit dies in them; i.e. they are becoming unconscious; another phrase for this is: *kanye isa asik balul*, the mind/spirit might go to the spirit village (in the high mountains)
kanye dilana	(Borrowed meaning, lit. thing to see images with) mirror, binoculars
kanye dob-	To lie, to deceive
kanye ib-	The spirit/mind is blocked; to be in despair
kanye kiklib-	To compare, give an example
kanye kunuk	To turn over in one's mind, to be troubled, to think of something with impatient desire
kanye kwinibum	The flying/floating of the spirits of the dead
kanye laklamak	They spread out their minds; become unconscious
kanye lelik ub-	To feel derision, malicious pleasure
kanye lobrob-	To be enchanted
kanye lukuldan-	To be excited, happy
kanye 'lyulamle	The mind becomes rotten; one becomes unconscious
kanye malye unmanil	The mind becomes bad in me; I am sad, I suffer
kanye mamun	The mind/soul is by itself, is satisfied
kanye mantalak	The mind is sucked out (in coma)
kanye meib-	To let the mind roam/fly; to falsely accuse somebody
kanye monokolongon	Depressed mood, to be silent, sad (also *kanye monokuk-*)
kanye morobrob-	To be startled
kanye sukub-	The mind cramps; to think of somebody with envy or jealousy
kanye teleb boubnil	My mind is filled with good; I am happy
farabrob kanye	The mind of someone who is fast and efficient in doing things, who gives away good things to others; energy
kulubkulub kanye	Mind directed towards sexual intercourse (in this way, each meaningful compound word can be formed)

(*continued*)

Table 6.1. *Continued*

lin kanye	The mind of someone who can act alone, behaves purposefully; plan, design
lon kanye	A freeing mind, the forgiving of injustice
otenen kanye	The mind of someone who only thinks of him/herself, egoism (That is) their mind, their business
tamubmalul kanye	A constantly turning mind; someone who does things his/her own solitary way
wa kanye	Mind directed towards the garden and gardenwork, diligence, assiduity

From Heeschen and Schiefenhövel 1983, p. 114 ff.

danger of being maltreated, harmed or killed, one can rely on food, shelter, assistance and protection, especially from one's clansmen. But also the genealogical lines of their mothers are very well represented in the minds of the Eipo. Talking to them about people in other villages, one can't but be struck by the precision with which some informants can trace family ties of virtually everyone 2 days walk away — almost the end of the world, as one does not venture farther away in the already mentioned and well-justified fear of being killed where one no longer has any relatives. It was we foreigners who roamed around in an effort to find the boundaries of the Mek culture; our hosts would not have dared to undertake such long journeys. Now, with acculturation having set in, many of them have travelled to the coast and other faraway places.

Knowing who is related to whom and, especially, who is related to oneself, was thus an extremely important element in Eipo society. While collecting demographic and genealogical data, one of the classic ethnographic tasks after arrival, we were impressed by the fact that virtually all adults were able to recall the names, clans and villages of their great-grandparents (see Schiefenhövel 1988). Some of the informants could present those kinds of data for a historic depth of five generations, a remarkable feat in an illiterate society. As long as A knows that B is, however distantly and indirectly, related to oneself, one can feel safe, one will be given shelter, food and, if needed, protection. The ties of blood are strong, no matter how thin they may be.

The importance of kinship, in traditional societies infinitely more important than in societies like our own, is reflected in complex kinship terminologies, a core element of social anthropology. These systems exhibit complex and sophisticated concepts of relatedness and relationships, most of which, regardless of whether one deals with one of the many potentially polygynous, the monogamous and the very rare polyandrous societies worldwide, centre around the reproductive unit of woman, man and their offspring. Tables 6.2 and 6.3 present the terms for relatives in two Melanesian societies, i.e. words or concepts for or of people who are connected by either descent or marriage ties.

Table 6.2. Eipo kinship terminology

	fa	*eli* (term of reference) *ni* (term of address, also term of reference for 2nd person singular)	Woman speaking	sihu	*lakanye* (term of reference) *nakanye* (term of address, also term of reference for 2nd person singular)
			Man speaking	brwi	*kil*, also: *u*
	mo	*elin* (term of reference) *nin* (term of address, also term of reference for 2nd person singular)	Woman speaking	brwi	*base kil*
				hubr	*lakanye* (term of reference) *nakanye* (term of address, also term of reference for 2nd person singular) also: *sal*
Man speaking	br	older: *dunye* younger: *weitnye* general term for brother; *siknang susuk* (when two brothers) *sik weicape* (when three and more brothers)		hu	*lakanye* (term of reference) *nakanye* (term of address, also term of reference for 2nd person singular) also: *sal*
Woman speaking	br	older: *dunye* younger: *weitnye* general term for brother: *makalnang*		wi	*kil*
				brwifa	*aupe*
				brwimo	*yamal kil*, *dukul*
				wifa	*ka*
				wimo	*yamal kil*
				wibrso	*ka*
Man speaking	si	older: *dukil*, *dukul* younger: *weit kil* general term for sister: *makal kil*		mosi	*nin kele* (term of reference) *nin* (term of address)
				brso	*me* woman speaking also: *cape*
				brdaso	*me*

(*continued*)

Table 6.2. *Continued*

Woman speaking	si	older: *dukul*, *dukil* younger: *weit kil* general term for sister: *siknang susuk* (when two sisters) *sik weicape* (when three and more sisters)		brsoso brdada brsoda mofasi brsoso	*me* *kilme* *kilme* *u* (term of reference) *nu* (term of address, also term of reference for 2nd person singular) also: *nin*, *elin* *me* woman speaking also: *cape*
Determination according to birth order among brothers and sisters:					
First-born, male: *dunye*, *fermale*: *dukil*				fabr	*ni kele*
Born in the middle (male and female): *nakaybye*				fabrwi	*nin*
noitamnye (all brothers and sisters having both older and younger brothers and sisters)			Man speaking	fabrda	*makal kil*
			Woman speaking	fabrda	*weit kil*, *dukil* (according to time of own birth)
Last-born (male and female):		*kelasirya*			
	so	*me* (singular)	Man speaking	fabrso	*weitnye*, *dunye*
		mape (plural)	Woman speaking	fabrso	*makalnang*
	da	*kilme* (singular)		siso	*me*
		kilmape (plural)		sida	*kilme*
	soso	*me*		mobr	*mam*
	soda	*kilme*		mosihu	*ni* (possibly only if same clan as speaker)
	daso	*me*			
	dada	*kilme*		mobrwi	*dukul* possibly also: *u*
	fafa	*aupe*		husiso	*me*
	famo	*u* (term of reference) *nu* (term of address, also term of reference for 2nd person singular)		husida wbrwi	*kilme* *u* (possibly within certain clan groupings only)
	mofa	*aupe*		husihu	*aupe* (possibly within certain clan groupings only)
	momo	*u* (term of reference) *nu* (term of address, also term of reference for 2nd person singular)		mobrso	*mam*

Table 6.2. *Continued*

	fafafa	*aupe kisok*
	famofa	*aupe kisok*
	famomo	*u kisok*
	fafamo	*u kisok*
	mofafa	*aupe kisok*
	momofa	*aupe kisok*
	mofamo	*u kisok*
	momomo	*u kisok*
	fasi	*cape kil* (also: *dukul*)
Man speaking	fasihu	*ka*, *base*
Woman speaking	fasihu	*lakanye* (term of reference)
		nakanye (term of address, also term of reference for 2nd person singular)
Man speaking	brda	*kilme*
Woman speaking	brda	*weit*, also *cape*
	wbrda	*kil*
Man speaking	sihu	*base*, *ka*
	mobrda	*nin kele* (term of reference)
		nin (term of address)
	fasiso	*me*
	fasida	*kilme*
	mofabr	*aupe*
	mofabrda	*nin kele* (term of reference)
		nin (term of address)
	mofabrso	*mam*
	fafabr	*aupe*

The terms for relationships within the *yala* (clan) have priority over other terms, i.e. to express a more distant relationship of one person to another one mostly uses the *yala* term, even though the consangineous relationship might require a different term.
Abbreviations: fa, father; mo, mother; br, brother; si, sister; so, son; da, daughter; hu, husband; wi, wife.
From Heeschen and Schiefenhövel 1983, p. 15–17.

Table 6.3. Trobriand kinship terms

Kinship terms	
Tabu(gu)	Grandparent, grandchild; father's sister, father's sister's daughter
Ina(gu)	Mother, mother's sister; mother's clanswoman
Tama(gu)	Father, father's brother; father's clansman; father's sister's son
Kada(gu)	Mother's brother and, reciprocally, sister's son and sister's daugher
Lu(gu)ta	Sister (man speaking), brother (woman speaking); woman of same clan and generation (man speaking), man of same clan and generation (woman speaking)
Tuwa(gu)	Elder brother (man speaking), elder sister (woman speaking); clansman of same generation but older (man speaking), clanswoman of same generation but older (woman speaking)
Bwada(gu)	Younger brother (man speaking), younger sister (woman speaking); clansman of same generation but younger (man speaking), clanswoman of same generation but younger (woman speaking)
Latu(gu)	Child, male or female
Marriage relationships	
(Ulo)mwala	Husband
(Ulo)kwava	Wife
Relationships-in-law	
Yawa(gu)	Father-in-law, mother-in-law
Lubou(gu)	Wife's brother, sister's husband
Iva(gu)ta	Husband's sister, brother's wife
Tuwa(gu)	Wife's elder sister, husband's elder brother
Bwada(gu)	Wife's younger sister, husband's younger brother

Adapted from Malinowski 1929/1968, p. 434.

To bear these manifold relationships in the mind and, especially, to know what a specific relative is entitled to by ways of proper address (an important issue of etiquette, telling of the importance these ties have in traditional societies), and by way of food and other treatment, is a challenge almost everybody, even children and juveniles, can cope with. It is essential that interactions flow smoothly when it comes to meeting one of the many relatives; with many of them an individual will have direct and meaningful dealings within his/her lifetime. Traditional etiquette, formalising relationships, duties, expectations, etc., plus the individual minds of interaction partners involved, emotional reactions, consciously thought-out strategies and enphronesis, must all come together well to ensure a good outcome for oneself. Most of the people one meets on a day-to-day basis in the close-knit communities of present-day traditional Melanesia, and in the neolithic past of all humankind, will have some kind of formal tie to oneself. This renders encounters and interactions much more predictable than in our anonymous societies, but there is of course still ample room for individually chosen attitudes and behaviours. Traditions governing interactions with a wide range of relatives create a kind of

framework of predictability, in which enphronesis tasks can be performed with less error and better efficiency.

BIG-MEN, GRANDS HOMMES — *MASTERS OF ENPHRONESIS IN NEW GUINEA HIGHLAND SOCIETIES*

The typical Papuan societies of New Guinea are 'acephalic', i.e. they do not have a system of hereditary chieftainship. Initiative and leadership, without which no society can function well in the long run, is exerted by men whose personalities make them natural leaders. They are usually in the prime of their lives, approximately 30–60 years old, vital, often but not always physically impressive, in any case intelligent, socially competent, self-assured, good orators who are able to convince others and draw them into the schemes they are pursuing (cf. Strathern 1971; Godelier 1982; Godelier and Strathern 1991).

These are the ones who perform and take politically meaningful decisions, such as when and where to build a new liana bridge across a roaring mountain river, how to solve legal problems, how to arrange large ceremonial feasts, and when and how to attack an enemy group with violent warfare. In the word's proper sense, these societies are patriarchal (*arche*, Greek for 'rule'), because virtually all important decisions affecting the society as a whole are taken by these 'big-men'.

It is an interesting question why these and other traditional societies in Melanesia and the rest of the world do not leave their fate in the hands of people, men and women, who would follow a hypothetical, archaic model of democracy and where mandates would shift from time to time. In the matrilinear Austronesian, i.e. non-Papuan, society of the Trobriand Islanders, political decisions are taken by powerful male chiefs, who get their position through inheritance via the matriline and after a process of discussion among other powerful men; the difference from the big-men system is that (a) they inherit their positions and (b) they keep them until they die.

In Papuan cultures, big-men cling to their positions as long as they can and forge all kinds of clever alliances to defend their place in the hierarchy. Political power rests in the minds and hands of men who can convince the others that they are able to do the job of political leadership — a true and archaic meritocracy. Big-men pursue a mix of group-orientated and personal goals; in ideal cases, the two will be largely or partially congruent. Apart from getting the group ready for intragroup fights and intergroup war (among the Eipo the war leader was a different man from the ceremonial and social leaders, even though he could also exercise power outside the battlefield) typical big-men activities are non-belligerent but nonetheless highly aspiring. Leaders in the wide and comparatively very populous valleys in the interior of New Guinea, such as the Wahgi valley and the Enga region in Papua New Guinea, and the Balim valley and westerly adjacent Dani areas in the Province of Papua in

West-New Guinea, are the organisers of outstandingly ambitious large-scale festive ceremonies, at which enormous numbers of pigs (often many hundred) and a variety of other very precious items, such as shells, polished stones, plaited bands, etc., are distributed.

Big-men utilise these carefully planned occasions as a stage on which they can perform the culminating acts of the ceremony in full sight of everybody — powerful and vulnerable, no job for cowards. Strathen (1971), working in the Hagen region of Papua New Guinea, stated:

> In Hagen [the] notion that the giver [the organizer] of a m*oka* is in some way superior, is explicit. To attain superiority he must give more than he received from his partner at the last bout of exchanges... It is the increment, strictly, which can be referred to as *moka*... (Strathern 1971, p. 10; brackets inserted by the present author).

Figure 6.1 lists some of the pathways involved in giving *moka* pigs; it becomes obvious, even from this schematic sketch (which cannot render the complexity of real-life transactions) how demanding things can be. In order to achieve success, the big-man must conduct the splendour of presentations as a breathtaking drama, casting a spell on everybody present, creating an unforgettable spectacle — and, no less important, he is able to pay back all the loans he has incurred during his daring 'financial' transactions — a big-man must position himself in the middle of the extremely complex web of gifts, debts and other transactions, both those connected to his own scheme and others occurring at other stages (Figure 6.2, op. cit.). He must have all the actual facts and figures and the numerous possible moves of his creditors, friends and enemies in his mind, mentally neatly organised and easily accessed, and he must not panic. Enphronesis is what he needs: he had better not assume a big-man's role if he can't perform and deliver. The big-man role must carry a sizeable sociobiological bonus — why otherwise would he run all the very real risks (shame, loss of face, perhaps even death if a fight is sparked by angry would-be recipients)? The risk war leaders take is even bigger.

CONCLUSION: DOES ENPHRONESIS PAY?

Strathern (1971, p. 12) states that the *moka*-exchange is rather a zero-sum game than a potlatch-type effort to dominate, and that a concept of balance is inherent in the Hagen exchange system. Notwithstanding this analysis, there is a tremendously competitive element involved (as is true also for the exchange ceremonies of Austronesian peoples, such as the *kula* among the Trobriand Islanders and similar systems elsewhere in Melanesia). These dramatised public ceremonies provide an immensely efficient mechanism for the protagonists to

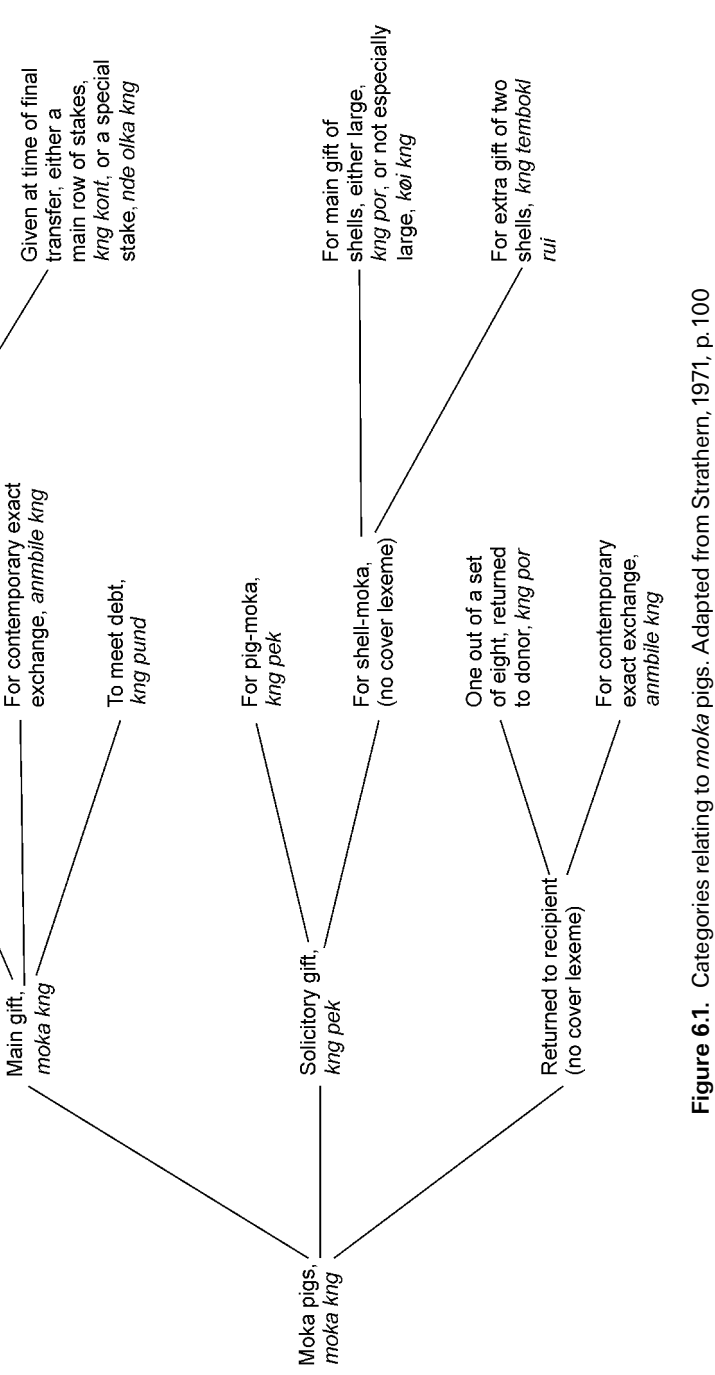


Figure 6.1. Categories relating to *moka* pigs. Adapted from Strathern, 1971, p. 100

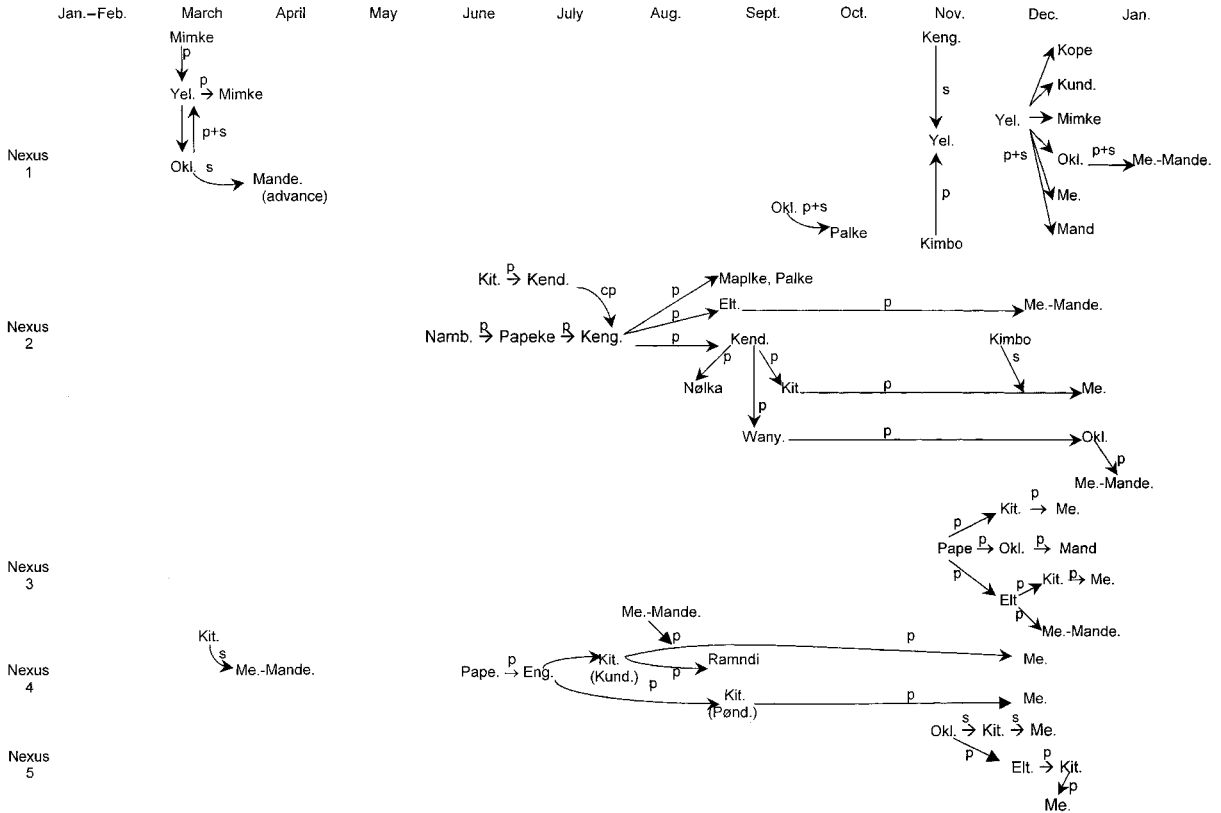


Figure 6.2. Chart of monthly moka activities, January 1964–1965. cp, cooked pig; p, pigs; p+s, pigs and shells; s, shells. Adapted from Strathern, 1971, p. 125

Table 6.4. Marriages of men of Kawelka Kundmbo clan

Numbers of marriages made by men of the clan group, up to the end of 1965

No. of marriages	Big-men		Adjutants	Ordinary men and others who have been married
	Major	Minor		
6	1	—	—	—
4	1	1	—	1
3	1	—	—	—
2	1	—	1	9
1	—	—	1	41
	4	1	2	51

Total: 58

NOTE: the ordinary man who had been married four times but was in 1965 without a current wife had a big-man as his sponsor, but seemed to be temperamentally incapable of retaining a marital partner.

Total number of wives of men at any time

No. of wives	Big-men		Adjutants	Ordinary men and others
	Major	Minor		
3	3	—	—	—
2	1	—	—	7
1	—	1	2	44
	4	1	2	51

Total: 58

Adapted from Strathern, 1971, p. 203.

be in the limelight and in the centre of talks long after the occasion — limelight praise as the ultimate reward for a risk-taking business?

Table 6.4 provides a clear answer, sociobiological in essence, even though Strathern (1971) was not phrasing his findings in this paradigm at the time of writing: big-men have many more wives than all the others, including their adjutants. In the Hagen case, due to the very careful data collection of the author, we have proof that all the risk taking, all the alpha-male stuff, can pay off in evolutionary terms: big-men simply have more (official) mating opportunities and are most likely to translate this into number of offspring.

In the small-scale Eipo society we have not been able to show clear inclusive fitness advantages of big-men vs. 'ordinary citizens', but it can be demonstrated (Schiefenhövel 1988, p. 80) that of 131 Eipo men, 36 (27.5%) were unmarried in August 1975, whereas of 128 marriageable women only one (!) had no

husband. This imbalance is mainly due to preferential female infanticide, which tremendously skews the sex ratio in favour of men (to approximately 150/100 for some cohorts), so that there are far from enough women to whom to get married. This problem is aggravated by the fact that 12% of the men were polygynous, taking extra women out of the pool. About one-quarter of the Eipo male population will have to live without a wife. Marriages are formally sealed and spouses are expected to have no extramarital sex (which, however, happens quite often).

Long-term bachelors (some of whom may have a chance to marry a widow later in life) can only hope to have a short-term love affair and thereby father children. Usually there is violent defence of the spouse by the rightful husband; most intragroup deaths are due to this (Schiefenhövel 2001). Yet, by far the majority of children are likely to come from married men. Eipo women, who are in a perfect position for female choice, can select among them. The majority of marriages are love marriages, despite the fact that parents tend to arrange partners in the cross-cousin mode. Most of the women select men who are physically able, intelligent, socially competent, hard and diligent workers and good fighters; in sum, men who are respected in the village. Enphronesis is involved in four of these categories. Long-term Eipo bachelors usually did not score well in them.

The Hagen society, with its comparatively large number of people and very advanced system of big-men roles, demonstrates more clearly than that of the Eipo (whose presentation ceremonies are less elaborate than those in the Hagen) that the quest for a high position in the hierarchy is no nonsense game. The stakes are high: one may lose face, but one may have many more children than those who do not dare to enter the highly competitive life of big-men alpha males. Many if not all societies [hunter-gatherers seem to have a more levelling ideal (see Wiessner 1996), perhaps curtailing the chances for individual males to stick out as much as New Guinea big-men] have similar cultural institutions which provide stages and open avenues for powerful men to publicly perform, thus advertising their good genes on the female choice market. Guts and enphronesis must have been strong selection factors in human history.

REFERENCES

Brothers, L. (1990). The social brain: a project for integrating primate behavior and neurophysiology in a new domain. *Concepts Neurosci.* **1**, 27–51.

de Waal, F. (1982). *Chimpanzee Politics*. Jonathan Cape, London.

Godelier, M. (1982). *La Production des Grands Hommes. Puvoir et Domination Masculine chez les Baruya de Nouvelle Guinée*. Fayard, Paris.

Godelier, M. and Strathern, M. (eds) (1991). *Big Men and Great Men. Personifications of Power in Melanesia*. Cambridge University Press, Cambridge, and Editions de la Maison des Sciences de l'Homme, Paris.

Heeschen, V. and Schiefenhövel, W. (1983). *Wörterbuch der Eipo-Sprache: Eipo–Deutsch–Englisch*. Reimer, Berlin.

Laporan Expedisi Ilmiah Lembah X. 2 Okt. sampai dengan 20 Des. (1969) di Irian Barat. Komando Daerah Militer XVII 'Cenderawasih'. Mimeo, Jayapura.

Malinowski, B. (1929/1968). *The sexual life of Savages in Northwestern Melanesia*. Routledge and Kegan Paul, London.

Premack, D. and Woodruff, G. (1978). Does the chimpanzee have a 'theory of mind'? *Behav. Brain Sci.* **4**, 515–526.

Saulnier, T. (1960). *Les Papous, Coupeurs de Têtes. 167 Jours dans la Préhistoire*. Editions du Pont Royal, Paris.

Schiefenhövel, W. (1976). Die Eipo-Leute des Berglands von Indonesisch-Neuguinea: Kurzer Überblick über den Lebensraum und seine Menschen. Einführung zu den Eipo-Filmen des Humanethologischen Filmarchivs der Max-Planck-Gesellschaft. *Homo* **26** (4), 263–275.

Schiefenhövel, W. (1988). *Geburtsverhalten und reproduktive Strategien der Eipo — Ergebnisse humanethologischer und ethnomedizinischer Untersuchungen im zentralen Bergland von Irian Jaya (West-Neuguinea), Indonesien*. Reimer, Berlin.

Schiefenhövel, W. (1991). Eipo. In T.E. Hays, (ed.), *Encyclopedia of World Cultures*, *vol. II, Oceania*. G.K. Hall & Co, Boston, MA, pp. 55–59.

Schiefenhövel, W. (1999). Modern models of the past: the life of the Eipo, Highlands of West New Guinea. In H. Ullrich (ed.), *Hominid Evolution. Lifestyles and Survival Strategies*. Edition Archaea, Gelsenkirchen, 268–276.

Schiefenhövel, W. (2001). Kampf, Krieg und Versöhnung bei den Eipo im Bergland von West-Neuguinea — Zur Evolutionsbiologie und Kulturanthropologie aggressiven Verhaltens. In W. Fikentscher (ed.), *Begegnung und Konflikt — eine kulturanthropologische Bestandsaufnahme*. Bayerische Akademie der Wissenschaften, Philosophisch-Historische Klasse, Abhandlungen, Neue Folge, Heft 120. C.H. Beck, München, pp. 169–186.

Strathern, A. (1971). *The Rope of Moka. Big-men and Ceremonial Exchange in Mount Hagen*. Cambridge University Press, London, New York.

Wiessner P. (1996). Levelling the hunter: constraints of the status quest in foraging societies. In P. Wiessner and W. Schiefenhövel (eds), *Food and the Status Quest. An Interdisciplinary Perspective*. Berghahn, Providence, Oxford, pp. 171–191.

PART III

Developmental Aspects of the 'Social Brain'

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

7

Big Brains, Slow Development and Social Complexity: the Developmental and Evolutionary Origins of Social Cognition

DAVID F. BJORKLUND[1] AND JESSE M. BERING[2]

[1]Department of Psychology, Florida Atlantic University, Boca Raton and
[2]Department of Psychology, University of Arkansas, Fayetteville, USA

Homo sapiens has been deemed distinctive from the rest of the animal world on the basis of a number of dimensions. Humans are (presumably) the only species to use language; although others use tools, only humans make tools to make other tools, and no other species comes close to achieving the technological prowess that humans do; other species, such as chimpanzees (e.g. Whiten *et al.* 1999), pass on traditions from one generation to the next, one definition of culture, but, again, chimpanzee cultures pale by comparison with the diversity and complexity of human cultures. And humans' general learning abilities, freed from the apparent constraints of 'instincts', surpass those of other animals (although many species, from butterflies to blue jays, display some domain-specific learning skills that put corresponding human abilities to shame). Although we do not wish to dispute our species' special accomplishments in areas related to language, technology, culture and learning, what has set humans apart from the rest of the animal kingdom, we argue, is something that they share with many other species: social intelligence. The course of human evolution is not a straight line from the common ancestor we shared with chimpanzees, dating back 5–7 million years ago, to modern times, and there is,

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

of course, no single factor responsible for human cognitive evolution; other selection pressures, such as those related to changes in technology, diet and climate (e.g. Potts 1998; Kaplan *et al*. 2000; Calvin 2002), surely contributed to shaping the modern human mind. But a number of psychologists and anthropologists have argued that the single most important factor in human cognitive evolution was not the advent of language or the development of tools, but the increasing complexity of hominid social groups (e.g. Humphrey 1976; Byrne and Whiten 1988; Alexander 1989; Dunbar 1992, 2001; Bjorklund and Harnishfeger 1995; Geary and Flinn 2001). From this perspective, our remarkable technological skills have been co-opted from the 'intelligence' evolved to deal with cooperating, competing and understanding conspecifics.

Social cognition — cognition about social relationships and social phenomena — is not a unitary skill, but rather can be thought of as a set of hierarchically-arranged, relatively specific abilities evolved to deal with the variety of social problems faced by our ancestors (Geary 1998; Bugental 2000). Following Geary, social cognition can be divided into *social-individual modules*, which include processing information related to non-verbal behaviour, language, facial processing and theory of mind, and *social-group modules*, including processing information related to kin, ingroup membership, outgroup membership and social ideologies. Despite sharing this modular view of cognitive abilities with mainstream evolutionary psychology (e.g. Tooby and Cosmides 1992; Buss 1995; Cosmides and Tooby 2002), we also believe that some aspects of social cognition are relatively domain-general in nature (see Bjorklund and Pellegrini 2002; Corballis 2002), and that changes in foundational processes, such as speed of processing, inhibition abilities or working memory capacity have contributed to enhanced social cognitive skills across domains.

Regardless of one's belief about the primacy of social intelligence in human cognitive evolution or the degree to which such processes are domain-specific or general in nature, social cognition undisputedly plays a central role in everyday human functioning and is critical for success in any human group. Moreover, social intelligence, as all other forms of intelligence, develops, and it also developed in our ancestors. In this chapter, we argue that the confluence of three factors was necessary for the evolution of human social cognition: a large brain, an extended juvenile period, and a socially complex environment. In support of this argument, we review research examining various aspects of the development of social cognition in children and great apes, particularly chimpanzees, the latter of which, we argue, can provide some insight to what the social-cognitive abilities of our ancient ancestors may have been like. We conclude by examining the role of developmental plasticity (afforded in humans by a large brain and an extended juvenile period) in human ontogeny and specifically in human phylogeny, suggesting that looking at how species-atypical rearing environments affect the social cognition of great apes may

serve as a window to how developmental plasticity may have influenced the evolution of the human mind.

BIG BRAINS, EXTENDED YOUTH AND SOCIAL COMPLEXITY

It would seem that having a big brain, relative to one's body size, would be all that is needed to have increased intelligence. After all, that's what brains are for. But big brains are not equally likely to be developed in all species. For example, big brains imply an ability to learn. Learning would seem always to be a good thing, but it has its limitations. An animal that is able to learn many different contingencies, the location of many different sources of food or shelter and the identities of hundreds or thousands of conspecifics, will be able to put that information to good use only if it lives a relatively long life. Brains are metabolically expensive (Aiello and Wheeler 1995), and a short-lived animal would have little use of such vast knowledge and would do better to devote its limited time to finding sustenance, mating, and protecting its young rather than 'learning'.

Humans, of course, do have an extended lifespan as well as big brains, conditions necessary for the evolution of intelligence (as conventionally understood by psychologists and lay people alike). Relative to other mammals, humans' brains are far larger than expected for their body size (see Jerison 1973, 2002; Rilling and Insel 1999). For example, the encephalisation quotient (EQ) reflects the brain weight 'expected' for an animal for its body weight (Jerison 1973). Given the typical pattern of changes in brain and body weight in mammals, brain weight should increase at a certain rate relative to increases in body weight. Animals that have 'more brain' than expected, given their body size, have EQs greater than 1.0, whereas those with 'less brain' relative to their body size have EQs less than 1.0. The presumption is that animals with higher EQs can use their 'extra' brain for learning and related cognitive feats (Deacon 1997; Jerison 2002). This is an oversimplification, of course, as recognised by Jerison (2002), for it ignores potentially important details about species differences in the organisation of the brain. Nevertheless, it provides a reasonable and convenient metric for assessing the general 'learning and cognitive capacities' of different species, both living and dead.

Using Jerison's (1973) formula, modern chimpanzees have an EQ of about 2.3. The EQ for modern humans, however, is more than triple this, about 7.6 (Jerison 1973; Rilling and Insel 1999). Furthermore, if we look at fossil evidence, we find a regular increase in EQs over the course of hominid evolution, with the EQs of *Australopithecus afarensis*, dating back about 3.5 million years ago, being only slightly greater (about 3.1) than that of modern chimpanzees, with EQs increasing for *Homo habilis* (about 4.0; 2.5

million years ago), and *Homo erectus* (about 5.5; 1.5 million years ago) (Tobias 1987).

To achieve a brain as large as modern humans possess required modification in the timing of development. Human infants are born 'early', relative to other primates. That is, given their size and the ratio of brain to body weight, human infants, were they to follow the typical primate schedule, would be born after about 18–24 months of gestation (Gould 1977). But the head of such a neonate could not fit through the birth canal of its mother. As a result, human babies are born immature, and even at that, birth is often a difficult and dangerous process because of the large skull that must make its way through the birth canal.

But humans' brains continue to grow at a rapid rate following birth. Brains grow rapidly relative to body size for all mammals during the prenatal period, but this rate quickly declines after birth (Gould 1977; Bonner 1988). In contrast, human infants retain the rapid rate of brain growth well into the second year of life, afforded by a prolongation of the closure of the cranial sutures. Extending the time that the brain can grow results in the production of more neurons (Finlay and Darlington 1995; Finlay, Darlington and Nicastro 2001) and greater dendritic and synaptic growth, so that the human brain has both more neurons and more interconnections among neurons than the brains of other primates (Gibson 1991). Although most areas of the human brain show an increase in size relative to our chimpanzee cousins, some areas have increased more than others, suggesting that specific selective pressures were at work in shaping the hominid brain. Particularly important is the human neocortex, which is implicated in complex cognition and hypothesised to be the locus of important inhibitory control (e.g. Luria 1973; Fuster 1984) and, by the assessment of some scientists (e.g. Deacon 1997; Eccles 1989; Rilling and Insel 1999; Barton and Harvey 2000), is estimated to be about 200% the size expected for an ape of comparable body size (although others claim that, while large, it is proportional to the size of the overall brain; e.g. Smendeferi *et al.* 1997; Jerison 2002).

Homo sapiens's extended juvenile period provides the time both to grow their brains and to acquire information necessary for their survival. Although females in modern societies usually attain puberty in their early teen years, anthropologists have suggested that our female ancestors likely did not begin having children until their late teens or early 20s (e.g. Bogin 1999; Kaplan *et al.* 2000). Delaying reproduction can have dire consequences, suggesting that the advantages of such prolonged immaturity must have been substantial. Human culture and technology are diverse, making it necessary for children to have flexible intelligence to learn the specific rules and appropriate behaviour of their society. As we have commented previously, this takes not only a large brain to accomplish but also time. We argue that the requirement to deal with the social complexity of hominid cultures, in a big-brained, slow

developing animal, was a major selection pressure in the evolution of human intelligence.

In making this claim, we are not suggesting that humans are in any way 'special', or the inevitable product of evolution. Rather, the relationship among brain size, length of juvenile period and social complexity is a general one, reflecting an evolutionary pattern within primates in general, e.g. Dunbar (1992, 1995, 2001) has shown that brain size and social complexity are highly related among primates (correlation between size of neocortex and group size = 0.76). Also among primates, brain size is associated with the length of the juvenile period, with animals with larger adult brains (relative to body size) having longer juvenile periods (Bonner 1988). Based on these relationships, we argue that both large brains and delayed development are necessary to succeed in complex societies.

In support of this contention, Joffe (1997) compared aspects of brain size and structure with length of the prereproductive period and aspects of social complexity for 27 primates, including humans. Joffe measured the size of the non-visual neocortex, the part of the primate brain associated with complex problem solving and memory, and reported that the proportion of the lifespan spent as a juvenile was positively correlated with group size *and* the relative size of the non-visual neocortex. Correlation does not imply causality, of course, and, in fact, we believe that no single causal link between these three factors (brain size, length of juvenile period, social complexity) is possible. The three factors surely interacted synergistically, with large brains and an extended juvenile period being necessary for mastering the ways of one's group, and social complexity in turn exerting selection pressures for increased brain size and an extension of the juvenile period.

Other factors besides social complexity surely also were involved in this synergistic relationship. For example, Kaplan *et al.* (2000) proposed that big brains and an extended juvenile period were required to master the demands associated with a shift to a more complex diet, including hunting and the collection of difficult-to-obtain roots and tubers. Anthropological evidence indicates that such a shift did, indeed, occur about 1.9 million years ago with our *Homo erectus* ancestors (Wrangham 2001). Kaplan *et al.* (2000) note that one major difference between modern chimpanzees and humans living in traditional groups is their diet. Although both hunt, chimpanzees acquire relatively few calories from meat; likewise, chimpanzees get many of their calories from easy to acquire ripe fruit, whereas modern hunter/gatherers depend more on difficult to obtain (and prepare) roots and tubers. Kaplan *et al.* note that whereas young children, similar to chimpanzees, can learn to gather fruit, it takes many years before the skills associated with hunting and successful gathering are attained. Although these observations may seem to constitute an alternative explanation for human cognitive evolution (i.e. alternative to the hypothesis advocating social complexity as a driving force),

we see it as complementary. The food-collection and processing skills of modern humans are passed from one generation to the next via mechanisms of social learning, and, we argue, it is only in socially complex environments that such skills could be acquired. And it is to an examination of such skills, both in human children and in great apes, that we now turn.

SOCIAL COGNITION IN CHILDREN AND IN APES

CONTINUITY OF MENTAL FUNCTIONING

Human social-cognitive evolution (as the evolution of any characteristic) was based on an extension of the traits possessed by their ancestors. Phylogeny provides constraints on what can (and can likely) evolve and what cannot (or cannot likely) evolve. Although any new characteristic (physical, behavioural or cognitive) must pass through the sieve of natural selection, the basic materials of evolution are found in the phenotypes and genotypes of one's predecessors and in the constraints afforded by ontogeny (e.g. Raff 1996; Gould 2002). For humans, this means looking at the fossil record for hints of the cognitive abilities of our *Homo* and australopithecine ancestors (e.g. Mithen 1996; Wynn in press) and examining modern-day great apes, who last shared a common ancestor with humans 5–7 million years ago. Humans, of course, did not evolve from chimpanzees or any other living great ape (bonobos, gorillas or orangutans), with each contemporary species undergoing millions of years of evolution distinct from that of the others. But these genetic cousins represent the best guess of what our common ancestor may have been like, and examining their behaviour and cognition can give us a glimpse at how the human mind may have come to be.[1]

In many ways, human beings are much like chimpanzees (*Pan troglodytes*). Both humans and chimpanzees live in socially complex groups, both use tools, pass on information from one generation to another via social learning, have large brains relative to body size and an extended juvenile period (McGrew

[1] Both common chimpanzees (*Pan troglodytes*) and bonobos, or pygmy chimpanzees (*Pan paniscus*) are equally related to humans, with these two species diverging from one another only about 1 million years ago (Byrne 2002). The two apes share many similarities, but also differences [e.g. chimpanzees hunt, make tools, have elaborate customs, are dominated by male coalitions; bonobos are more socially cohesive, engage in frequent non-reproductive sex, and have more egalitarian roles between the sexes (e.g. Byrne 2002; de Waal 1997), although others have proposed that the differences are not as great as have been suggested (e.g. Stanford 1998)]. Which ape best reflects what our common ancestor may have been like is a matter of great debate. Unfortunately, bonobos have only recently received serious scientific study, whereas research on common chimpanzees, both in the wild and in captivity, is substantial (see Byrne 1995; Tomasello and Call 1997; de Waal 2001). As such, most evidence about what our common ancestor may have been like is inferred from research with chimpanzees, coupled with what evidence can be obtained from modern hunter/gatherers and the fossil record.

1992; Allman 1999; Bogin 1999; Whiten *et al.* 1999; Bjorklund and Pellegrini 2002). In many respects, some of the characteristics of modern humans can be seen as extensions of patterns displayed among social primates. Our similarity with chimpanzees should not be too surprising, given that we share an estimated 99% of our DNA with this species.

By emphasising the continuity of mental abilities among species within a phylogenetic clade, we do not mean to imply that all cognitive changes are quantitative in nature, or simply matters of degree. It is almost certain that some differences between the cognition of modern humans and our ape ancestors were qualitative in nature, such as those reflecting humans' unique language ability. Nonetheless, such discontinuous changes reflect reorganizations of existing abilities/brain tissue or the emergence of new abilities/brain tissue from existing stock, as opposed to the evolution of totally 'new' skills.

In the sections below we examine several aspects of social learning in children and great apes, looking also, briefly, at factors that may have contributed to the evolution of such abilities in humans. Included in our review are various forms of social learning, theory of mind, and social reasoning.

SOCIAL LEARNING

The basis of social cognition is *social learning*, which refers to the acquisition of social information and behaviour or, more precisely, to 'situations in which one individual comes to behave similarly to others' (Boesch and Tomasello 1998, p. 598). Higher-level cognition is not necessary for social learning to occur, although more and increasingly varied information can be transmitted between individuals when more sophisticated types of social learning (such as imitation) are involved. Both human children (e.g. Bandura 1986) and chimpanzees (e.g. Whiten *et al.* 1999) acquire information via mechanisms of social learning. In fact, there seems little debate about the fact that chimpanzees are marvellous social learners. Rather, debate surrounds the mechanisms of social learning used by human children and great apes and the cognitive abilities that underlie them.

Types of Social Learning

Perhaps the most potent form of social learning, at least among humans, is direct teaching. Human parents universally instruct their children in ways great and small, from moulding their toddlers' fingers to handle a fork or chopsticks, to providing verbal directions to wait for the 'walk' sign to change before crossing the street. The phenomenon is so ubiquitous that no empirical investigation is necessary to prove the point. The situation is different, however, for chimpanzees. Despite thousands of hours of observation of chimpanzees, both in the wild and in captivity, there has been only a handful documenting

teaching (e.g. Boesch 1991, 1993; Greenfield *et al*. 2000). The most convincing evidence of teaching in chimpanzees has been provided by Boesch (1991, 1993), who observed female chimpanzees in Ivory Coast's Tai Forest showing their infants how to crack nuts. Nut cracking is found in only a few chimpanzee populations and is transmitted from one generation to the next, qualifying as an element of 'culture' (Whiten *et al*. 1999). It involves placing a nut on a rock and hitting it with another rock to open it. It is a skill that takes years to master (McGrew 1992) and is performed mainly by females, often in the company of their infants. On several occasions, Boesch observed female chimps positioning the nut and the anvil and hammer rocks so that all an infant had to do was strike the nut to open it, something that he never observed females without infants to do. At other times, he observed mother chimpanzees executing the actions especially slowly in the presence of their infants. These are impressive observations and consistent with the position that chimpanzees may occasionally directly teach their infants complex behaviours. However, the frequency with which such behaviour has been observed is low, and it appears not to be a common way in which social behaviour (with learning to use tools being a case of social behaviour) is acquired (see Bering 2001).

Perhaps the most obvious form of social learning, in humans and non-humans alike, is *observational learning* (Bandura 1986), which, as its name denotes, refers to learning behaviours simply by observing. Although one is tempted to assume that observational learning is accomplished by imitation, this is only one of several social learning mechanisms, and a rather sophisticated one at that. Comparative psychologists have identified several different social-learning mechanisms, with imitation being one of the more sophisticated. (Child developmental researchers have usually assumed that any new behaviour acquired by children by observation qualified as imitation, primarily because they were less concerned with the underlying mechanisms involved than they were with other factors, such as memory or the content of what was learned; see Want and Harris in press.) For example, social learning can occur via *local enhancement* (Thorpe 1956), in which an animal notices that interesting things are occurring at a particular location (e.g. tasty bugs running out from under a log), is drawn to that location, and, in a process of trial and error, discovers a useful behaviour (e.g. lifting a dead log often reveals a meal of delectable insects). *Stimulus enhancement* (Spence 1937) is a similar process, with an animal attracted not so much to location, but by the particular features of stimulus (e.g. a log). More sophisticated mechanisms include *mimicry*, the duplication of a behavior without any understanding of the goal of that behaviour, and the cognitively more sophisticated processes of *emulation* and true *imitation*.

Emulation (sometimes referred to as *goal emulation*; Whiten and Hamm 1992) refers to comprehending the goal of a model and engaging in similar behaviour to achieve that goal, without necessarily replicating the specific

actions of the model (Boesch and Tomasello 1998; Tomasello 1996, 2000), e.g. chimpanzee A may observe chimpanzee B picking up and moving a log to reveal a nest of ants, which are then eaten; chimpanzee A may then find another log, which she rolls back and forth and eventually is rewarded with a meal of ants. Unlike in mimicry (or true imitation), the chimp does not replicate the behaviours of the model but rather, through a trial and error process, achieves the desired goal she had observed another animal attain.

In contrast to emulation, true imitation, according to some theorists, requires the observer to take the perspective of the model, to understand the model's goal, and to replicate important portions of the model's behaviour (Tomasello, Kruger and Ratner 1993; Tomasello 1996, 2000; Boesch and Tomasello 1998). From this viewpoint, true imitation is not the mindless matching behaviour of an unsophisticated organism, but reflects rather sophisticated cognitive processing.

The questions of interest here concern the ontogeny of social learning in children and in great apes. Are the social-learning abilities observed in human children merely more sophisticated versions of those observed in apes, particularly chimpanzees, or do apes and children approach social-learning contexts in substantially different ways?

Social Learning in Infants and Young Children

Neonatal imitation, in which newborns match the facial behaviour of a model, usually tongue protrusion, has been repeatedly demonstrated in human neonates (e.g. Meltzoff and Moore 1977, 1992; Vinter 1986; for review, see Anisfeld 1991). In these experiments, infants match the behaviour of a model, such as tongue protrusion. Thus, nothing new is learned in such encounters, and the underlying mechanism is best described as mimicry (there is debate about whether this behaviour is related to later imitation in infancy or is better thought of as a fixed-action pattern, designed to facilitate nursing or communication between the newborn and its mother, e.g. Jacobson 1979; Bjorklund 1987; Legerstee 1991).

Infants' abilities to copy the behaviours of others improves over the first 2 years of life (see Piaget 1962; Uzgiris and Hunt 1975), although most research has focused on the replication of vocal or motor behaviours or simple actions on objects (e.g. shaking or touching objects; see Want and Harris in press). Want and Harris (in press) have recently reviewed research on children's social-learning mechanisms and concluded that infant and preschool children frequently engage in mimicry and perhaps imitation, but rarely develop an understanding for the affordances of a tool and generalise its use to other contexts (a form of emulation; but see apparent exception from Bauer and Kleinknecht in press). Several researchers have shown that infants will reproduce faithfully the actions of a model, including irrelevant and

non-functional ones (such as hand waving) when adults demonstrate some behaviour on objects to a child (e.g. dropping a marble into a tube; e.g. Harnick 1978; Sibulkin and Uzgiris 1978; Nagell, Olguin and Tomasello 1993; Whiten *et al*. 1996). For example, in a study by Nagell, Olguin and Tomasello (1993), an adult experimenter modelled one of two behaviours to 2 year-old children using a rake to retrieve an out-of-reach object. The children copied the actions of the adults, even when a more efficient way of solving the problem was possible. In another experiment, Whiten *et al*. (1996) showed 2–4 year-old children a clear box with a desirable object inside. The box contained a series of bolts and a latch, and an experimenter demonstrated several ways (of many) in which the box could be opened and the reward obtained. Children tended to copy the exact actions of the model, even though other, more direct routes to opening the box were available. They also copied non-functional or redundant actions.

These, and other data, suggest that young children are prone to mimicry. However, from such findings it is not clear that this is because of their limited cognitive systems (i.e. they are incapable of emulation or true imitation) or because of social demand characteristics (i.e. imitating another is socially rewarding in its own right). For example, the robust tendency of infants and toddlers to copy the often arbitrary behaviours of an unfamiliar experimenter has made the use of deferred-imitation tasks in infancy and early childhood a window to memory development (see Bauer 1997 for review). Such matching behaviour may be important, not only for acquiring new knowledge and skills but also for facilitating social relations.

Other research shows quite convincingly that infants and toddlers are able to understand the goal of a model. For instance, several studies have shown that toddlers will copy the *intended* actions of a model, even when those actions were not successfully completed. In one study (Meltzoff 1995), 18 month-old children observed adults perform actions on objects. Sometimes the actions succeeded (e.g. picking up a dumbbell-shaped object, pulling on the wooden cube on one end and removing the cube) and sometimes they did not (e.g. the model's hand slips off the end of the cube, failing to remove it from the dumbbell). Relative to children in a control condition who did not see a demonstration of the dumbbell, children in both the successful and unsuccessful conditions removed the cube on the end of the dumbbell when given the opportunity. Children appeared to realise what the model *intended* to do to achieve an inferred (but not witnessed) goal. In a second experiment, 14 and 18 month-old infants watched as either a person or a mechanical device acted on an object (e.g. a person removing the cube end of a dumbbell or a vice-like machine pulling on one end of the dumbbell to remove the cube). Infants who watched the person were twice as likely to copy the actions than infants who watched the mechanical device, suggesting that by 14 months of age, infants understand that people (but not inanimate objects) have intentions

(goals) that are sometimes worthy of imitating (see also Carpenter, Akhatar and Tomasello 1998).

As noted previously, little child development research has focused on the mechanisms underlying young children's observational learning, particularly with respect to tool use (Want and Harris in press). A study by Want and Harris (2001) is an exception. They evaluated 2 and 3 year-old children's ability to learn to retrieve a treat from a hollow tube using a stick. If the stick were placed in one end of the tube and used to push the treat, the treat fell out the opposite end and children received a reward; however, if the stick were placed in the other end of the tube, the treat was pushed through a hole in the tube and was lost to the child. When children were shown both a correct and an incorrect solution to the problems, 3 year-olds were later able to imitate the actions to retrieve the treat successfully. Two year-old children, in contrast, did not fully understand the relationship between the model's actions with the tool and the outcome, but simply copied the behaviour of the model (mimicry), resulting in chance performance.

Social Learning in Great Apes

There is both direct and indirect evidence of social learning in chimpanzees. Indirect evidence comes from the documentation of wild chimpanzees' transmission from one generation to another of complex, learned behaviours, including fishing for ants and termites, nut cracking and styles of grooming (Whiten *et al.* 1999). Whiten and colleagues identified a total of 39 different behaviours that varied considerably from one troop to another, making any explanation of transmission other than by social means untenable.

Direct evidence comes from laboratory studies of chimpanzees using a variety of research formats and a range of socially acquired behaviours (see Galef 1988; Whiten and Ham 1992; Custance, Whiten and Bard 1995; Whiten 1996; Parker and McKinney 1999). There is good evidence that chimpanzees are able to copy arbitrary actions, such as facial expressions or hand signs (e.g. Hayes and Hayes 1952; Sanders 1985; Custance, Whiten and Bard 1995), much as infants and young children will copy the gestures of a model. It is less clear, however, what means chimpanzees use in social learning when objects such as tools are involved. For example, in the study by Nagell and colleagues (1993) described earlier, in which children were observed to reproduce faithfully the behaviours of a model in retrieving a desired object (interpreted as mimicry), a different pattern was obtained when chimpanzees served as subjects. Unlike the children, the chimpanzees seemed to ignore the precise actions of the model, but rather used a single inflexible strategy for retrieving the object with the rake (see similar results for chimpanzees by Tomasello *et al.* 1987). Similarly, in the study by Whiten *et al.* (1996) in which 2–4 year-old children reproduced the actions (both relevant and irrelevant) of a model to open a clear box containing

a treat, chimpanzees again generally failed to copy the behaviours of the model, but nonetheless learned to open the box using combinations of observed behaviours. The most parsimonious interpretation of these findings is that chimpanzees acquired the behaviour by emulation, understanding and eventually achieving the goal, but *not* doing so by explicitly matching their behaviour to that of a model. Emulation and local enhancement have been suggested as principal mechanisms underlying the 'cultural transmission' of nut cracking observed in a group of West African chimpanzees (Boesch and Tomasello 1998). The same argument could be made for the cultural transmission of termite and ant fishing in other chimpanzee populations (e.g. Goodall 1986).

Other studies have reported greater evidence of imitation (as opposed to other forms of social learning) in chimpanzees. In an experiment with a similar apparatus to that used in the Whiten *et al.* (1996) study, chimpanzees witnessed a model perform specific sequences of actions to open the box (e.g. open bolt 1, open bolt 2, rotate pin, turn handle) over a series of trials (Whiten 1998). Although two of the chimpanzees opened the box on the first trial and three on the second, there was no evidence that the animals matched the sequence of behaviours displayed by the model on these early trials. Greater evidence of matching the sequences was found on the third trial, although even the four animals that successfully opened the box on this trial did not copy the particular *behaviours* within those sequences with great fidelity. In related research, in which a chimpanzee trained to use a tool to acquire food from a tube served as a model for naïve chimpanzees, 3 and 4 year-old animals (but not 2 year-olds) in the modelling condition subsequently were successful in solving the food-retrieval problem and in generalising their behaviour to a more difficult task (Bard, Fragaszy and Visalberghi 1995). Other researchers using a similar paradigm, however, report that chimpanzees that observed the model solve a 'honey fishing' problem successfully, acquired the skill no faster than animals not witnessing a successful trial (Hirata and Morimura 2000).

The best evidence for true imitation in chimpanzees comes from studies with human-reared, or *enculturated*, animals (Tomasello, Savage-Rumbaugh and Kramer 1993; Bering, Bjorklund and Ragan 2000; Bjorklund, Bering and Ragan 2000; Bjorklund *et al.* 2002). In these studies, following a baseline period in which the animals explored the target objects, a model displayed some specific actions on the objects (e.g. holding cymbals by their handles and striking them together to produce noise), and then either immediately (Tomasello, Savage-Rumbaugh and Kruger 1993), after a 10-min delay (e.g. Bering, Bjorklund and Ragan 2000; Bjorklund *et al.* 2002), or a 24 or 48 hour delay (Tomasello, Savage-Rumbaugh and Kruger 1993), the objects were returned to the animal. Each of these studies reported that reproduction of the observed behaviour was significantly greater than at baseline and, in the Tomasello *et al.* study, significantly greater than that observed for mother-raised chimpanzees. Thus,

these studies suggest that in addition to emulation, imitation of actions on objects is within the ability of great apes, at least under some rearing conditions.[2] We will have more to say about the possible effects of enculturation on the social-cognitive development of great apes later.

Given the complexity of the data and the diversity of methods and findings, it is not surprising that there is not a consensus concerning the social-learning abilities of chimpanzees and the other great apes. It is clear that chimpanzees learn much from social situations but apparently, under most circumstances, do so via the mechanisms of local enhancement or emulation. Emulation should not be thought of as an unsophisticated mechanism, for in many situations it may be more useful to understand the goal and the affordances of the objects involved, and to explore different ways to solve a problem, than to copy a behaviour blindly (mimicry). Unlike chimpanzees, human infants and young children will copy the behaviour of a model, even if that behaviour has little or no relevance to solving a problem; such mimicry serves as much to maintain and facilitate social interaction as it does to acquire information. But older infants and toddlers also seem to understand the *intention* of a model (e.g. Meltzoff 1995), indicating a capacity for true imitation. Interestingly, much before the age of 3 or 4 years, children seem less capable of engaging in emulation (Want and Harris in press), although task factors apparently affect their performance (Bauer and Kleinknecht in press).

THEORY OF MIND

According to some theorists (e.g. Tomasello 1999; Tomasello, Kruger and Ratner 1993), some forms of social learning (e.g. true imitation) require understanding the intentions of other individuals. Put differently, the learner must understand what the model has in mind in executing some action and evaluate the episode accordingly. Such skills have been generally studied under the rubric *theory of mind*, which refers to the tendency to construe other individuals and their behaviours in terms of mind-related constructs, such as beliefs, desires and intentions (Wellman 1990; Perner 1991; Lillard 2002). Our interactions with others are based upon what Wellman (1990) has termed *belief-desire reasoning*: we understand that our own behaviour, and assume that the behaviour of others, is based upon what we believe (know, expect) and

[2] Of the three enculturated chimpanzees in the Tomasello, Savage-Rumbaugh and Kruger (1993) study, one was a common chimpanzee (*Pan troglodytes*) and two were bonobos (*Pan paniscus*). The studies by Bering, Bjorklund and Ragan (2000) and Bjorklund *et al*. (2000) included enculturated orangutans (*Pongo pygmaes*) as well as enculturated common chimpanzees. Although the orangutans in the Bering, Bjorklund and Ragan and Bjorklund *et al*. studies also showed signs of imitation, there is reason to believe that they do not engage in the same strategies of social learning as chimpanzees (see Call 1999).

what we desire (want, wish), and that our beliefs and desires can sometimes be different from those of others.

Although social learning is certainly possible without possessing a theory of mind, cooperating, competing, deceiving and negotiating with others is clearly facilitated when one understands that others have wants, desires and knowledge that is sometimes similar and sometimes different from one's own. This insight has led cognitive developmental scientists to believe that theory of mind is the basis of children's social cognition: it is difficult to imagine a person succeeding in any human culture without such understanding, and its development has been the most researched topic in child development since the first scientific study with children was published (Wimmer and Perner 1983).

However, the initial experimental research on theory of mind predates Wimmer and Perner's study by 5 years, and was performed by Premack and Woodruff (1978) with chimpanzees. As with social learning, a large and varied literature has arisen on theory of mind in great apes, and the opinions of whether apes (particularly chimpanzees) have a theory of mind or do not is equally diverse and controversial (e.g. Whiten 1996; Heyes 1998; Povinelli, Bering and Giambrone 2000; Suddendorf and Whiten 2001; Byrne 2002; Povinelli and Bering 2002). Our position and, we believe, the position of most researchers in the child development field (e.g. Flavell and Miller 1998; Lillard 2002) is that theory of mind is not an all-or-none thing but shows gradations, beginning, perhaps, with self-recognition and progressing through the more complicated forms of secondary representations (e.g. John knows that Claudia knows that he likes Monika). In the following sections, we review briefly aspects of theory of mind development in children and chimpanzees and examine some of the factors associated with its development in children, and at least one factor (increasing inhibitory control) that may have played an important role in the evolution of social cognition in the line that led to modern humans.

The Development of Theory of Mind in Children

Research previously cited demonstrating that 14 month-old infants will copy the intended (although not executed) actions of a human (but not mechanical) model (e.g. Meltzoff 1995), suggests that the rudiments of theory of mind (understanding that others have intentions) are rooted early in life. But more complicated theory-of-mind abilities would seem to require a clear distinction between self and others. One way of assessing whether children possess a cognitive sense of self is the *mirror self-recognition test*. Beginning at about 18 months of age, children who, unbeknownst to them, have a mark placed on their foreheads, will touch that mark when they subsequently look in a mirror. Children much younger than this typically touch the reflection in the mirror,

seemingly not realising that the images in the mirror are themselves (Lewis and Brooks-Gunn 1979). Children are a bit slower to recognise themselves in pictures and videos under similar conditions (e.g. Povinelli, Landau and Perilloux 1996; Povinelli and Simon 1998), suggesting that the sense of self develops gradually over the preschool years.

Another early sign of theory of mind, differentiating between what a child prefers and what another person prefers, was illustrated in a study by Repacholi and Gopnik (1997), who gave 14 and 18 month-old toddlers choices of two types of food to eat, Pepperidge Farm goldfish and raw vegetables. Children then watched as a woman tasted the two types of food and showed a preference opposite to the one the child had displayed (e.g. showing disgust for goldfish and pleasure for vegetables when the child's preferences were the reverse). Placing her hands between the two foods, the woman then said, 'Can you give me some?' The 14 month-olds gave the woman the food *they* liked, whereas the 18 month-old children gave the woman the food that *she* liked, apparently recognising that their likes and dislikes were different from those of another person.

Children's theory of mind has been most frequently investigated on variants of *false-belief tasks*. For example, in a widely employed task, used in the pioneering study by Wimmer and Perner (1983), children watch as a treat is hidden in a specific location (e.g. in a box). Another person (Maxi) is present when the treat is hidden but then leaves the room, at which time the treat is moved to a new location as the children watch. The children are then asked where Maxi will look for the treat when he returns. A robust finding is that most 4 year-old children can solve the problem, stating that Maxi will look where the treat was originally hidden, whereas most younger children state that Maxi will look for the treat in the new hiding place, apparently not realising that Maxi's knowledge is different from their own (e.g. for reviews, see Wellman 1990; Perner 1991; Flavell and Miller 1998; Lillard 2002).

Another form of theory of mind task reveals the importance of understanding that the goals of another person sometimes conflict with one's own. Peskin (1992) showed preschool children a series of stickers, some more attractive than others. She then introduced 'Mean Monkey', a hand puppet controlled by the experimenter, who would play a game with the children. Mean Monkey would ask the children which of the stickers they really wanted and which stickers they did not want, and then select the children's favourite sticker, leaving them with the least desirable ones. By 4 years of age, children understood the dynamics of the interchange and quickly learned to tell Mean Monkey the opposite of their true desires. Younger children rarely caught on and played most of the game telling Mean Monkey the truth and not getting the stickers they wanted.

Children's understanding of the intentions of others becomes more complex, of course, as children develop the ability to use second-order representations

involving recursive relations (e.g. she knows, that I know that she knows . . .), which are often the basis of sophisticated social interactions as well as the plots for novels and soap operas (e.g. Chandler and Lalonde 1996).

There is debate about the extent to which theory of mind and related social cognition reflects domain-specific abilities vs. a more general set of representation skills, developing over the preschool years. For example, Baron-Cohen (1995) has proposed a series of four domain-specific modules: an *intentionality detector* (ID), permitting the inference that a moving object may have some intention toward the individual; an *eye-direction detector* (EDD) that interprets eye gazes from other individuals; *shared-attention mechanisms* (SAM), involving triadic interactions between two people and an object (e.g. child, mother and toy); and a *theory of mind mechanism* (ToMM), similar to belief–desire reasoning described earlier. Some of these abilities may be found in all animals (e.g. ID), whereas others may be unique to humans (e.g. ToMM). In children, they develop over the preschool years and the more advanced modules (e.g. SAM and ToM) seemingly do not come on line for people with autism (e.g. Baron-Cohen, Leslie and Frith 1985; Perner *et al*. 1989).

In contrast to this domain-specific view, other researchers argue that the underlying representational abilities involved in theory of mind (particularly those associated with passing false-belief tasks) are more general in nature, and are correlated with the acquisition of other forms of non-social cognition (e.g. Gopnik and Astington 1988), such as passing appearance/reality distinction tasks (e.g. Flavell, Green and Flavell 1986), and can be predicted from the development of more general working memory and executive function tasks (e.g. Carlson, Moses and Nix 1998; Perner, Stummer and Lang 1999; Perner and Lang 2000; Carlson and Moses 2001). Our own position is that both domain-specific and domain-general abilities likely underlie the development of children's theory of mind skills, and that, regardless of their cognitive specificity, are influenced by social/environmental factors occurring over childhood.

Theory of Mind in Chimpanzees

Given chimpanzees' impressive social-learning ability, it has often been assumed that they possess some rudiments of theory of mind. Despite more than 25 years of research since the issue was first put to an experimental test (Premack and Woodruff 1978), there is not a consensus on the 'mindreading' abilities of chimpanzees. Chimpanzees and orangutans (but not monkeys) do 'pass' the mirror self-recognition tasks (e.g. Gallup 1979; Miles 1994), although at older ages than shown by children (Amsterdam 1972; see also Povinelli 1995). Also there have been numerous observations of deception by chimpanzees and other great apes (e.g. Whiten and Byrne 1988), which, on the surface, would seem to require an understanding that others have

knowledge and desires different from one's own. For example, Belle was shown the hidden location of food and would then lead a small groups of chimps to it (Menzel 1974). However, when Rock, a dominant male chimp, was present, he would take the food for himself. As a result, Belle would wait until Rock left before retrieving the food, and on other occasions would lead the troop in the opposite direction of the food, doubling back later to the cache.

This and other equally impressive signs of deception reveal a complex and flexible cognitive system, but not necessarily one that requires theory of mind. In most cases of deception in great apes, such as the one involving Belle, the animal could have learned a specific response in a specific situation (e.g. I lose the food when I retrieve it when Rock is around). In these contexts, it may not be the case that an ape is reading the mind of another, but only that a particular behaviour is successful in a particular situation (see Bjorklund and Kipp 2002).

Conducting false-belief tasks with non-verbal chimpanzees can be difficult to both execute and interpret, but a variety of experiments assessing the components of false belief have been conducted with mixed evidence. In a complicated non-verbal false-belief task (not solved by children until age 5), food was hidden behind a barrier, outside of the apes' view but within the view of a human 'communicator', who helped the apes locate the treat by placing a marker on the correct container. The communicator sometimes made mistakes, which the apes learned to ignore (i.e. the communicators sometimes placed the marker on a container in which the ape knew the treat was not hidden). Once the apes were able to perform these tasks, the false-belief portion of the experiment began. The communicator watched as the treat was hidden in one of the containers and then left the room. A second person then switched the location of the treat, while the ape watched. When the communicator returned, she placed the mark where she had seen it hidden previously. None of the two orangutans or five chimpanzees passed the task. They seemed not to realise that the communicator had a false belief about the location of the treat, and they continued to select erroneously the container marked by the communicator (Call and Tomasello 1999).

In research somewhat analogous to the 'Mean Monkey' experiment of Peskin with preschool children, Boysen and Berntson (1995) placed two chimpanzees on opposite sides of a partition, with one ape acting as the 'selector' — who got to choose between a larger and smaller array — and the other as the 'observer' — who actually received whichever array the selector pointed to. Whichever pile of treats was *not* selected went to the selector. The rub, of course, was that if the selector wanted the larger portion all to itself, it would have to inhibit its response toward that array and instead point to the smaller one. This did not happen. The attractiveness of the food rewards interfered with task performance such that the selectors rarely seemed to 'get it': they would go for the larger array on almost all trials. Despite the

substantial learning abilities of these chimpanzees in other contexts, they, like the 3 year-old children in Peskin's (1992) study, were not able to inhibit their response to select the desirable prize, and as a result never received it. (However, when the actual food items were removed from the scene and replaced by Arabic numerals reflecting the quantity of each respective array, the selector immediately adopted the effective reward strategy, selecting the smaller number and thus receiving the larger cache of food. These animals, that had previously been trained to associate Arabic numerals with specific quantities, thus possessed the underlying ability to inhibit their behaviour under appropriate conditions, but did so only in a highly unusual, species-atypical context.)

Other research brings into question whether chimpanzees understand that looking at something implies knowledge. For example, in a series of experiments, Povinelli and Eddy (1996) taught chimpanzees to make a reaching response to a human to receive a food treat. In the test conditions, two caretakers stood in front of a chimpanzee, one who could see the chimp and gave it the treat when the ape reached to her, and one who could not see the chimp because her eyes were in some way occluded (e.g. eyes closed, blindfolded, bucket placed over her head, back facing the chimp). Chimpanzees rarely made reaching responses to someone who had her back to them, but otherwise they rarely discriminated between the caretaker who could see and the one who could not. They also responded randomly in a condition in which both caretakers had their backs to them, but one was looking over her shoulder at the ape (see also Reaux, Theall and Povinelli 1999). These findings suggest that although chimpanzees understand that a person with her back towards them cannot see them (which is often, but not always, correct), they apparently fail to understand that it is the eyes that are responsible for acquiring useful knowledge about the immediate environment.

Recent experiments using a more naturalistic context have suggested that, in some situations, chimpanzees may indeed understand that seeing can lead to knowing. In a series of experiments (Hare *et al.* 2000; Hare, Call and Tomasello 2001), a dominant and a subordinate chimpanzee were housed in connected cages, with food placed at various positions in the cages. When the food was placed so that both animals could see it, the dominant chimpanzee always got the food. But when the food was placed so that only the subordinate animal could see it, it successfully obtained the food on most trials, waiting until the dominant chimp was not looking before retrieving the treat. These findings suggest that, at least under some conditions, chimpanzees know what other chimpanzees can and cannot see and make decisions based on that knowledge.

As we've noted, there is great debate about the extent to which great apes in general and chimpanzees in particular, possess theory of mind. However, although our bias is to believe that chimpanzees do not possess social-cognitive

abilities comparable to those of 4 year-old children, they do display behaviours that may serve as the basis for higher-level cognitive abilities, and if similar skills were possessed by our common ancestor, this provides ample raw material for natural selection to produce the social-cognitive skills characteristic of modern humans. Some of the social-cognitive abilities of chimpanzees likely reflect unique adaptations, evolved since splitting from the line that led to *Homo sapiens*, and may therefore be unrelated to human abilities. Others, however, particularly to the extent that they are shared in greater or lesser degree with bonobos, gorillas and orangutans (e.g. Parker and McKinney 1999; Byrne 2002), were also likely possessed by the common ancestor last shared by both humans and chimpanzees, and thus may have served as the fodder for the evolution of the modern human mind.

Factors Influencing the Ontogeny of Theory of Mind

Theory of mind is not a unitary phenomenon and thus surely has multiple factors influencing its development. This is the case, we argue, regardless of whether important aspects of theory of mind turn out to be domain-general or domain-specific in nature. In this section we review briefly two general sets of factors that have been proposed to be causally related to theory of mind development in children: the social environment and executive functions.

Given that theory of mind serves to provide an understanding of social interactions, it should not be surprising that factors in the social environment may be related to its development. For example, preschool children's performance on false-belief tasks is positively related to the number of adults and older peers they interact with on a daily basis (Lewis *et al.* 1996). It is similarly positively related to family size (e.g. Perner, Ruffman and Leekam 1994; Jenkins and Astington 1996; but see Cutting and Dunn 1999 for an exception), particularly the number of older siblings a child has (Ruffman *et al.* 1998). One suggestion for the beneficial effect of older siblings and peers on theory of mind development is that they provide more opportunities for discussions of mental states, pretend play, managing social conflict, and reasoning about social issues, among others (Lewis *et al.* 1996; Ruffman *et al.* 1998; Smith 1998). For example, Ruffman *et al.* argued that having older siblings stimulates fantasy play, which helps children represent 'counterfactual states of affairs', a skill necessary for solving false-belief tasks. Similarly, Leslie (1987) has proposed that pretend play is an indicator of metarepresentational abilities, which serve to foster an understanding that someone else may represent things differently (have different knowledge or beliefs) from oneself. Consistent with this argument are research reports demonstrating positive relationships between the amount of cooperative social play preschoolers engage in (often with a sibling or parent) and later understanding of other peoples' feelings and beliefs (Astington and Jenkins 1995; Youngblood and

Dunn 1995). Cummins (1998a) provides a different interpretation of these findings, suggesting that siblings are frequent competitors for resources, with older siblings typically having the advantage because of their greater size and mental abilities. It is therefore to the younger siblings' advantage to develop whatever latent talents they have to aid them in their social competition with their older siblings, and developing 'mindreading' skills sooner rather than later should help them in such interactions.

A different set of influences, internal rather than external to the child, concerns what have been termed *executive functions* (e.g. Zelazo and Frye 1997; Perner, Stummer and Lang 1999; Perner and Lang 2000). Executive functions refer to cognitive abilities involved in planning, executing and inhibiting actions. Of the various components of executive function related to theory of mind, inhibition mechanisms have received the most attention (e.g. Russell *et al*. 1991; Perner, Stummer and Lang 1999; Carlson and Moses 2001; Bjorklund and Kipp 2002). Cognitive inhibition refers to the ability to inhibit certain thoughts and behaviours at specified times. Developmental differences in the ability to inhibit prepotent responses (behavioural and cognitive) have been proposed to play a major role on a wide range of cognitive and social tasks, including memory, selective attention, problem solving, reading, resisting temptation and delaying gratification, among others (see e.g. Bjorklund and Harnishfeger 1990; Dempster 1992; Harnishfeger 1995; Kochanska *et al*. 1996; Lehman *et al*. 1997; Lorsbach, Katz and Cupak 1998; Wilson and Kipp 1998).

With respect to theory of mind, many tasks require children to inhibit a dominant response to 'pass' the task, e.g. in the Peskin (1992) 'Mean Monkey' study, children had to resist telling the puppet which sticker they really wanted, which 3 year-olds had a difficult time doing. Similarly, in research by Russell and his colleagues (1991), 3 year-old children were shown a series of windows, some of which had treats in them. In order to get the treat, the children had to select the non-treat window. The children had a difficult time doing this, and repeatedly failed to get a treat, seemingly being unable to inhibit their 'pick-the-treat' response (cf. the behaviour of chimpanzees in the study by Boysen and Berntson 1995, discussed above). Several research projects have examined the relation between preschool and early school-age children's performance on theory of mind tasks and performance on batteries of executive function tasks, including inhibition (e.g. Perner, Stummer and Lang 1999; Carlson and Moses 2001). The studies generally find positive correlations between the two sets of tasks, usually in the range 0.30–0.60. These findings are consistent with the interpretation that a certain level of basic information-processing skills, specifically inhibition (although working memory skills are also separately implicated in theory of mind performance; see Moses, Carlson and Sabbagh 2002), is required to pass many theory of mind tasks. This line of reasoning also suggests that one or more domain-general skills are implicated in theory of mind development.

The Role of Inhibition in the Evolution of Social Cognition

In addition to playing a significant role in the ontogeny of social cognition, inhibition abilities have been proposed to play a critical role of the phylogeny of social-cognitive skills within the hominid line (e.g. Stenhouse 1974; Bjorklund and Harnishfeger 1995; Bjorklund and Kipp 1996, 2002). For example, Bjorklund and Kipp (1996, 2002) have suggested that the ability to inhibit sexual and aggressive responses was important for moderating behaviour in increasingly complex social settings. The ability to engage in deception, discussed briefly above with regard to chimpanzees, also requires the ability to inhibit a prepotent response, something that apes and young children often have a difficult time doing (e.g. Russell *et al.* 1991; Boysen and Berntson 1995). Greater inhibition abilities may also have been required for successful parenting. As children were born increasingly premature (compared to other great ape newborns) and were increasingly dependent on their parents (particularly their mothers) for nurturance, females may have needed greater inhibition abilities to deal with an often unruly and aversive offspring. Successful parenting in contemporary times often requires putting an infant's needs ahead of one's own, and this involves delaying one's own gratification, resisting distractions that may take one away from the infant, and inhibiting aggressive responses toward a sometimes difficult ward. The prolonged immaturity, associated with a big brain and social complexity, may have necessitated a mother who could better inhibit behaviour contradictory to the best interests of her infant. In fact, Bjorklund and Kipp (1996) hypothesised that ancestral women would have required greater inhibitory skills than men, in part because of the demands of parenting but also in part because of a need to mask interests in a potential mate until they had a greater opportunity to 'check him out' [this is because of the greater potential investment females have in sex than males (pregnancy and subsequent care of an offspring), making a more careful evaluation of a potential mate of greater importance for females than for males, following the tenets of parental investment theory (e.g. Trivers 1972; Bjorklund and Shackelford 1999)]. Consistent with this argument, there is some evidence that contemporary females exhibit greater inhibition abilities in the social and behavioural realms (but not the cognitive realm) than males, beginning in the preschool years (see Bjorklund and Kipp 1996; Stevenson and Williams 2000).

THE DEVELOPMENT OF SOCIAL REASONING

Theory of mind is the most studied form of social cognition during childhood. However, there are other forms of social cognition, perhaps dependent upon, but otherwise relatively independent of, theory of mind, which have also been investigated and which may have their origins in our evolutionary past. One

such set of skills includes those involved in social exchanges and the ability to detect people who may be breaking the social rules. Cosmides and Tooby (1992) summarised a series of experiments contrasting the logic people use to solve abstract problems with the same logic they use to solve *social-contract problems*. For the abstract problems, Cosmides and Tooby used variants of the Wason (1966) task. Adult are shown four cards, such as the ones displayed below:

A	G	2	7

Participants are given the following rule: 'If a card has a vowel on one side, then it must have an even number on the other side'. They are told that they must determine if the set of cards in front of them conforms to the rule or not, and should turn over the fewest number of cards possible to determine the truth of the rule. This is a difficult task, one which many college students fail (the correct answer is 'A' and '2'). But the task becomes easier when the same logic is applied to a social-contract problem. For example, adults are given the following cards:

Beer	Coke	16 years old	25 years old

They are then asked to test the following rule: 'If a person is drinking alcohol, then he/she must be at least 21 years old'. Now, most adults solve the problem easily, turning over the 'Beer' and '16 years old' cards, recognising immediately that what's on the other side of the 'Coke' card is irrelevant, as is what's on the other side of the '25 years old' card. Cosmides and Tooby (1992) proposed that the reason for the discrepancy in performance between the 'abstract' and 'social contract' versions of the same problem is that people do not use a general problem-solving ability to solve all logical problems, but rather have evolved domain-specific 'cheater detectors' that are limited to social contracts. From a similar perspective, the social-contract problems reflect *deontic reasoning*, which is reasoning about what one may, should or ought to do, whereas the abstract problems reflect *descriptive*, or *indicative* reasoning, which implies only a description of 'facts', and no violation of social rules.

Several developmental studies have investigated deontic reasoning in young children, presenting them with simplified variants of the Cosmides and Tooby task (e.g. Cummins 1996a; Harris and Nuñez 1996). For example, in a study by Harris and Nuñez (1996, Experiment 4) 3 and 4 year-old children were told short stories, some of which involved breaking a prescriptive rule (deontic condition) and others that had the same content but without breaking any rule (descriptive condition). For instance, in the deontic condition, children were told, 'One day Carol wants to do some painting. Her Mum says if she does some painting she should put her apron on'. Children in the descriptive

condition were told, 'One day Carol wants to do some painting. Carol says that if she does some painting she always puts her apron on.' Children were then shown four drawings, e.g. Carol painting with her apron, Carol painting without her apron, Carol not painting with her apron, and Carol not painting without her apron. Children in the deontic condition were then told, 'Show me the picture where Carol is doing something naughty and not doing what her Mum said'. Children in the descriptive condition were told, 'Show me the picture where Carol is doing something different and not doing what she said'. Both the 3 and 4 year-old children were more likely to select the correct picture in the deontic condition (72% and 83% for the 3 and 4 year olds, respectively) than in the descriptive condition (40% for both the 3 and 4 year-olds). Like adults, young children were better able to reason correctly about a problem in which a social contract was being violated than one in which no such social obligation was mentioned.

Similar to the arguments she made for the positive effects of having older siblings on the development of theory of mind (see above), Cummins (1996a, 1998a,b) has argued that children's deontic reasoning is innate and evolved in the context of dominance hierarchies within primate groups. It is important for social primates to know what someone of one's rank is permitted to do and not to do and to recognise when others are following or breaking the rules, which could have consequences for one's standing in the hierarchy. Cummins proposed that social (deontic) reasoning evolved from the combined effects of a large brain and a complex, hierarchically organised primate social system. We would add that an extended childhood was also likely necessary for deontic reasoning to evolve. And, counter to Cummins, we do not classify deontic reasoning as 'innate', but rather propose that children are predisposed to attend to and are sensitive to feedback related to social contracts/exchanges, which facilitates the development of deontic reasoning, and that one's position a social hierarchy influences rate of development of such reasoning.

THE ROLE OF DEVELOPMENTAL PLASTICITY IN COGNITIVE EVOLUTION

We have emphasised throughout this chapter that the confluence of a large brain, a prolonged juvenile period and social complexity set the stage for the evolution of the contemporary human mind. Although we have proposed that changes in underlying inhibitory control was one factor influencing changes in social-cognitive abilities in hominid evolution, we have not been specific about the mechanisms by which such changes may have taken place. Of course, as all contemporary evolutionists, we believe that mutations that afforded fortuitous advantages would have been selected and, given the niche in which our ancestors lived, selection would have favoured mutations resulting in larger

brains, slower development and enhanced social-cognitive skills. As we hope we have shown, our common ancestor with chimpanzees likely possessed substantial social-cognitive abilities, including a sensitivity to dominance hierarchies, complex social-learning skills, and possibly the ability to use deception in certain contexts; from such a base, relatively minor changes in social-cognitive functioning, if they should accumulate over many generations (and perhaps species), could yield the more advanced social-learning skills, including theory-of-mind abilities, characteristic of modern *Homo sapiens*. But we do not believe that evolutionary change proceeds by random mutational change alone; rather, factors associated with patterns of early development may also contribute to cognitive and behavioural evolution. More specifically, we argue, as have others (e.g. Gottlieb 1992, 1998; Ho 1998) that the plasticity of behaviour early in life can provide material for natural selection, and that such plasticity is greater for big-brained, slow-developing creatures such as humans, chimpanzees and their ancient predecessors.

PLASTICITY AND DEVELOPMENT

Mammals in general, and humans in particular, show substantial plasticity of cognition and behaviour early in development. It has long been known that social and learning-related behaviours of mammals, including rats, mice, monkeys and humans, are greatly influenced by early environment (e.g. Harlow 1959; Skeels 1966; Cairns 1979) and, counter to earlier claims (e.g. Spitz 1945; Harlow, Dodsworth and Harlow 1965), the effects of early experience can themselves be modified by later experience (e.g. Suomi and Harlow 1972; Clark and Clark 1976). For example, rhesus monkeys raised without the benefit of social contact (in terms of either conspecifics or humans) displayed aberrant social and sexual behaviour that remained relatively stable into adulthood (e.g. Harlow, Dodsworth and Harlow 1965). In a now classic study, Suomi and Harlow (1972) showed that placing the isolates, beginning at 6 months of age, into daily contact with a younger, socially inexperienced monkey resulted in subsequent normal social and sexual behaviour by the isolates when they were introduced into the monkey colony at 1 year of age.

The evidence of the reversibility, and thus plasticity, of behaviour in human children is even more impressive. For example, research dating back to the 1930s has demonstrated that children reared in stultifying institutions, or otherwise experiencing physical and social deprivation, demonstrate signs of social and cognitive retardation that persist to varying degrees into adolescence and beyond (e.g. Spitz 1945; Provence and Lipton 1962; Dennis 1973); yet, radical changes in living conditions can result in comparably radical changes in these children's social and intellectual development (e.g. Skeels 1966; Koluchova 1976; Clark and Hanisee 1982; O'Connor *et al*. and the English and Romanian Adoptees Study Team 2000). To take a recent example,

children living in Romanian orphanages, who experienced extreme social and nutritional deprivation and consequent physical, social and intellectual deficits, were adopted by English parents (O'Connor *et al.* 2000). As a result of such placement, children adopted within the first 2 years of life attained normal IQs by age 6. Children who were not adopted until 24–42 months of age also displayed IQs within the normal range (mean IQ = 90), but significantly lower than children adopted at earlier ages. This study, and others like it, demonstrates substantial plasticity of human intellectual functioning over childhood, although such plasticity decreases with age (although the late-adopted children did not attain as high an average IQ as the early-adopted children, it is possible that these children's IQs will increase as they spend more time in intellectually stimulating homes).

One reason for the impressive plasticity observed in human children, we believe, is their delayed development. Because brain development progresses relatively slowly in humans, experiences early in life can be more easily modified than if brain development were to proceed more rapidly. We know, of course, that new synapses are formed throughout life, not just in humans but also in other mammals, including rats (e.g. Greenough, Black and Wallace 1987). Actual growth of the human brain, however, including the process of myelination (e.g. Yakovlev and Lecours 1967), continues through the second decade of life, prolonging the time new skills can be easily acquired and the effects of 'old' experiences reversed.

PLASTICITY, DEVELOPMENT AND EVOLUTION

Epigenetic Theories of Evolution

Children inherit not only a species-typical genome, but also a species-typical environment (Gottlieb 1992, 1998, 2000; Lickliter 1996; Bjorklund and Pellegrini 2000, 2002; Oyama 2000). Although evolution is typically conceived as changes only in the genome, *epigenetic theories of evolution* hold that evolution is better conceived as reflecting changes in *developmental systems*, of which genes are an important part, but only a part (Gottlieb 1992; Ho 1998). A developmental systems perspective views both ontogeny and phylogeny as occurring as a result of the bidirectional interactions between all levels of organization related to an organism, from the genetic (DNA $\longleftrightarrow$ RNA $\longleftrightarrow$ proteins) through the cultural (culture $\longleftrightarrow$ behaviour). Structure (i.e. the organism itself, including its DNA, muscles, neurons) and function (i.e. activity emanating from the structures as well as events external to the organism) cannot be viewed as distinct, for development involves the continuous and bidirectional relation between structure and function at all levels of organisation (see Gottlieb 1991, 2000). From this perspective, changes in non-genetic aspects of a developmental system, such as the early

parent–infant environment, can contribute to evolutionary changes. Epigenetic theories of evolution view a developing organism's response to environmental changes as a mechanism for phylogenetic change. Natural selection still plays an important role in evolution, but it is the *developmental plasticity* of an organism that provides the creative force for evolution.

But how can changes in the behavioural phenotype of an animal as a result of early experience influence the evolution of behaviour and cognition? The inheritance of acquired characteristics has been appropriately rejected long ago, and we are not advocating here any mechanisms reminiscent of Lamarckian inheritance. Rather, consistent with what we are learning about the human genome (see International Human Genome Sequencing Consortium 2001; Venter *et al*. 2001), most DNA is inactive; 75% of the human genome is composed of intergenic (or 'junk') DNA. Gottlieb (1992) has suggested that radical changes in developmental conditions can result in the activation of inactive DNA (or the inactivation of active DNA), which can result in behavioural or even morphological changes, which in turn can be acted upon by natural selection. Similarly, novel environments may be fatal for some members of a species, whereas others, with extreme values of some alleles, may survive, thus changing substantially the distribution of genotypes (and phenotypes) in a population. Gottlieb (1992) has proposed that animals with large brains and substantial behavioural plasticity are more likely to adapt to novel environments than smaller-brained and less behaviourally flexible species, the implication being that larger-brained animals should show faster rates of evolution than smaller-brained animals (see also Wyles, Kunkel and Wilson 1983; McKinney 1998).

According to Wyles, Kunkel and Wilson (1983, p. 4396), animals that acquire new skills will use them 'to exploit the environment in a new way . . . [the] non-genetic propagation of new skills and mobility in large populations will accelerate anatomical evolution by increasing the rate at which anatomical mutants of potentially high fitness are exposed to selection in new contexts'.

There is limited evidence of behavioural changes in mammals carrying over two generations. For example, Denenberg and Rosenberg (1967) showed that female rats that had been handled as infants (removed daily from the home cage and placed in a tin can with shavings for 3 min) had grand-offspring who were more active and weighed less than the grand-offspring of non-handled rats under some conditions (e.g. when the second generation females had been exposed to a 'free' as opposed to a confined environment). Similarly, Ressler (1966) reported that the particular strain of foster grandparents influenced aspects of operant behaviour in mice. Although the nature of the mechanisms for such multi-generational effects were not known (unobserved influences on behaviour, physiology or milk content were proposed as possibilities), these

results suggested to Ressler (1966) that 'a non-genetic system of inheritance based upon transmission of parental influences is potentially available to all mammals (p. 267)'.

With respect to human evolution, we propose that cognitive-behavioural modifications in a hominid ancestor in response to novel environments may have contributed significantly to the social-cognitive abilities of *Homo sapiens*. Animals that benefited from an environmental change developed skills that enhanced their ability to deal with conspecifics (presumably from the activation of intergenic DNA, or because they possessed alleles, normally low in frequency in the population, that were now associated with survival). Assuming that the environmental changes were widespread and eventually stabilised, many animals would possess these skills, and important aspects of the species-typical developmental system would change.

The Enculturation Hypothesis

Although this proposal is necessarily speculative, aspects of it are amenable to empirical test. For example, using chimpanzees and other great apes as models for what our ancestors may have been like, changes in social-cognitive functioning can be evaluated as a function of changes in their early (and later) environments. We cannot know for certain the types of environmental changes our ancestors experienced (changes in diet and climate can be inferred from the geological record, but changes in parenting, for example, cannot be inferred so easily). One possibility, based on differences between modern humans and chimpanzees, is that changes in styles of mother–infant interaction contributed to social-cognitive modifications. As we mentioned previously, human mothers (and fathers) talk to their infants, instruct their infants and engage in triadic interaction, often using pointing gestures (referential communication) in the process. Chimpanzee mothers rarely engage in these behaviours, either in the wild or in captivity. What might be the consequences of rearing chimpanzees in a more human-like environment? Such species-atypical experiences in these large-brained, slow developing animals may result in cognitive abilities more child-like than found in mother-reared animals and provide evidence for the plausibility of the epigenetic evolutionary hypothesis proposed here.

As we mentioned earlier, several researchers have investigated aspects of this enculturation hypothesis (Call and Tomasello 1996). Some of our own research (e.g. Bering, Bjorklund and Ragan 2000; Bjorklund *et al.* 2002) and that of Tomasello and his colleagues (1993) have found support for this position for deferred imitation of actions on objects. Other researchers have observed, under conditions of experimental control, referential pointing, specifically the use of pointing to direct the attention of another individual, in enculturated chimpanzees and orangutans but not in nursery- or mother-reared animals (Povinelli, Nelson and Boysen 1992; Call and Tomasello 1994). Such research

indicates that species-atypical environments in human-reared great apes produce modified patterns of social cognition, more in line with that of human preschool children than is found in animals raised by their mothers.

We do not know the nature of these changes. Perhaps enculturation produces only molar (i.e. behavioural) levels of change in sociality, leading to enhanced imitative learning abilities (see Carpenter, Tomasello and Savage-Rumbaugh 1995; Povinelli 1996). For example, like human infants, enculturated chimpanzees will imitate the actions of a human model without any explicit reinforcement. Alternatively, enculturation may produce changes in the epigenetic system, leading to the phenotypic expression of cognitive abilities that are otherwise suppressed under natural conditions (see Tomasello 1990; Bjorklund and Pellegrini 2002).

We have heard the complaint that it is illogical to believe that enculturation studies can reveal latent talents in an animal. If the animal had this ability, it would only make sense that it used it in its natural habitat. Natural selection is conservative, the argument rightly goes, and animals do not possess abilities that are not used. These untapped abilities would be quickly eliminated from a species that made no use of them. For example, many subterranean species of vertebrates lose the function of their eyes when sight is no longer needed for survival. Random mutations resulting in the loss of vision are not selected against, leading to blindness in a relatively brief period of time (from a geological perspective).

Yet, a similar argument could be made about the effects of education on humans' cognitive abilities. Contemporary humans around the world engage in cognitive feats that their ancestors a mere 10,000 years ago could not have imagined. The most basic technological skills, such as reading and mathematics, were unknown to our ancient ancestors, as well as to most of our ancestors just several generations removed from the present. Yet, most 12 year-old children in literate societies can read proficiently and add, subtract, multiply and divide numbers with relative ease. These skills are not part of our basic hominid heritage. Language did not evolve to be read. They are what Geary (1995) has called *biologically secondary abilities*, which have not directly undergone selection pressure and are influenced by specific cultural practices. They are based on *biologically primary abilities*, which are universal and have undergone selection pressure over phylogeny. Language is an example of a biological primary ability, whereas reading is an example of a biologically secondary ability.

Geary discussed this distinction at length for humans, although others have hinted that this distinction might also be applicable to non-human mammals (e.g. Davis 1996). In particular, chimpanzees are also a big-brained, slow-developing species which, by some accounts, possess culture (e.g. Whiten *et al.* 1999). Although the differences in neural capacity and social complexity between humans and chimpanzees should not be underestimated, chimpanzees,

like humans, possess the mental plasticity to modify their cognition and behaviour to radically different environments and to develop novel cognitive abilities (biologically secondary abilities) in the process.

Findings from enculturation studies and all interpretations must remain tentative for now, in part because we do not know the specific aspects of human rearing that are important for these effects, and because of debate about the specific cognitive abilities that underlie performance (e.g. Povinelli 1996; Povinelli and Bering 2002). Nonetheless, the results of such studies suggest that developmental plasticity in a species likely not too dissimilar to our ancestors of 5–7 million years ago, could have played a role in social-cognitive evolution, and such plasticity may provide 'an experiential vehicle by which our hominid ancestors (using contemporary great apes as a model) could have begun to modify their cognition in the direction that resulted in *Homo sapiens*' (Bjorklund and Pellegrini 2002, p. 112).

CONCLUSION

Homo sapiens is an accomplished species. Humans' technological skills have permitted them, for better or worse, to dominate the world. Such skills are acquired in today's members of the species, and surely in yesterday's members, by means of social learning. As has been argued by others, human beings' unique technological intelligence evolved on the back of social intelligence. The need to cooperate, compete and generally understand fellow conspecifics was a driving force in the evolution of intelligence. But social complexity, such as that found in human groups the world over (and surely in our great ancestors) could not evolve under any set of circumstances. Substantial mental prowess, afforded by a large, slow-developing brain, which affords both the time to acquire the complex social and technological skills of human groups and the plasticity to modify behaviour in response to environmental changes, was a necessary condition for human social intelligence to evolve as it did. No one factor can be considered the 'cause' of the others; big brains, an extended juvenile period and social complexity surely evolved synchronously and in response to other environmental factors, such as changes in climate or diet. By examining social-cognitive abilities in children and in humans' closest genetic cousins, the great apes, we can garner insights about the underlying abilities that evolved and the primitive social intelligence on which subsequent evolution was based, and possibly about the mechanisms that led to the modern human mind.

AUTHORS' NOTE

This chapter was written while the first author was supported by a Research Award from the Alexander von Humboldt Foundation, Germany, and while

working at the University of Würzburg, Germany. We wish to express our appreciation to the Humboldt Foundation and to Wolfgang Schneider for their support of this work. Correspondence should be sent to David F. Bjorklund, Department of Psychology, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, USA; e-mail: dbjorklund@fau.edu.

REFERENCES

Aiello, L.C. and Wheeler, P. (1995). The expensive-tissue hypothesis: the brain and the digestive system in human and primate evolution. *Current Anthropology*, **36**, 199–221.

Alexander, R.D. (1989). Evolution of the human psyche. In P. Mellers and C. Stringer (eds), *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans*. Princeton University Press, Princeton, NJ, pp. 455–513.

Allman, J.M. (1999). *Evolving Brains*. Scientific American Library, New York.

Amsterdam, B. (1972). Mirror self-image reactions before age two. *Dev. Psychobiol.* **5**, 297–305.

Anisfeld, M. (1991). Neonatal imitation: a review. *Dev. Rev.* **11**, 60–97.

Astington, J.W. and Jenkins, J.M. (1995). Theory of mind development and social understanding. *Cogn. Emotion* **9**, 151–165.

Bandura, A. (1986). *Social Foundations of Thought and Action: a Social Cognitive Theory*. Prentice-Hall, Englewood Cliffs, NJ.

Bard, K.A., Fragaszy, D. and Visalberghi, E. (1995). Acquisition and comprehension of a tool-using behavior by young chimpanzees (*Pan troglodytes*): effects of age and modeling. *Int. J. Comp. Psychol.* **8**, 47–68.

Baron-Cohen, S. (1995). *Mindblindness: an Essay on Autism and Theory of Mind*. MIT Press, Cambridge, MA.

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

Barton, R.A. and Harvey, P.H. (2000). Mosaic evolution of brain structure in mammals. *Nature*, **405**, 1055–1058.

Bauer, P.J. (1997). Development of memory in early childhood. In N. Cowan (ed.), *The Development of Memory in Childhood*. Psychology Press, Hove, pp. 83–111.

Bauer, P.J. and Kleinknecht, E.E. (in press). To 'ape' or to emulate? Young children's use of both strategies in a single study. *Dev. Sci.*

Bering, J.M. (2001). Theistic percepts in other species: can chimpanzees represent the minds of non-natural agents? *J. Cogn. Cult.* **1**, 107–137.

Bering, J.M., Bjorklund, D.F. and Ragan, P. (2000). Deferred imitation of object-related actions in human-reared juvenile chimpanzees and orangutans. *Dev. Psychobiol.* **36**, 218–232.

Bjorklund, D.F. (1987). A note on neonatal imitation. *Dev. Rev.* **7**, 86–92.

Bjorklund, D.F., Bering, J. and Ragan, P. (2000). A two-year longitudinal study of deferred imitation of object manipulation in an enculturated juvenile chimpanzee (*Pan troglodytes*) and orangutan (*Pongo pygmaeus*). *Dev. Psychobiol.*, **37**, 229–237.

Bjorklund, D.F. and Harnishfeger, K.K. (1990). Children's strategies: Their definition and origins. In D.F. Bjorklund (ed.), *Children's strategies: Contemporary Views of Cognitive Development*. Erlbaum, Hillsdale, NJ, pp. 309–323.

Bjorklund, D.F. and Harnishfeger, K.K. (1995). The role of inhibition mechanisms in the evolution of human cognition and behavior. In F.N. Dempster and C.J. Brainerd

(eds), *New Perspectives on Interference and Inhibition in Cognition*. Academic Press, New York, pp. 141–173.
Bjorklund, D.F. and Kipp, K. (1996). Parental investment theory and gender differences in the evolution of inhibition mechanisms. *Psychol. Bull.* **120**, 163–188.
Bjorklund, D.F. and Kipp, K. (2002). Social cognition, inhibition, and theory of mind: the evolution of human intelligence. In R.J. Sternberg and J.C. Kaufman (eds), *The Evolution of Intelligence*. Erlbaum, Mahwah, NJ, pp. 27–53.
Bjorklund, D.F. and Pellegrini, A.D. (2000). Child development and evolutionary psychology. *Child Dev.* **71**, 1687–1708.
Bjorklund, D.F. and Pellegrini, A.D. (2002). *The Origins of Human Nature: Evolutionary Developmental Psychology*. American Psychological Association, Washington, DC.
Bjorklund, D.F. and Shackelford, T.K. (1999). Differences in parental investment contribute to important differences between men and women. *Curr. Dir. Psychol. Sci.*, **8**, 86–89.
Bjorklund, D.F., Yunger, J.L., Bering, J.M. and Ragan, P. (2002). The generalization of deferred imitation in enculturated chimpanzees (*Pan troglodytes*). *Anim. Cogn.* **5**, 49–58.
Boesch, C. (1991). Teaching among wild chimpanzees. *Anim. Behav.* **41**, 530–532.
Boesch, C. (1993). Toward a new image of culture in chimpanzees. *Behav. Brain Sci.* **16**, 514–515.
Boesch, C. and Tomasello, M. (1998). Chimpanzee and human cultures. *Curr. Anthropol.* **39**, 591–604.
Bogin, B. (1999). *Patterns of Human Growth*, 2nd edn. Cambridge University Press, Cambridge.
Bonner, J.T. (1988). *The Evolution of Complexity by Means of Natural Selection*. Princeton University Press, Princeton, NJ.
Boysen, S.T. and Berntson, G.G. (1995). Responses to quantity: perceptual vs. cognitive mechanisms in chimpanzees (*Pan troglodytes*). *J. Exp. Psychol. Anim. Behav. Proc.* **21**, 82–86.
Bugental, D.B. (2000). Acquisition of the algorithms of social life: a domain-based approach. *Psychol. Bull.* **126**, 187–219.
Buss, D.M. (1995). Evolutionary psychology. *Psychol. Inq.* **6**, 1–30.
Byrne, R. (1995). *The Thinking Ape: Evolutionary Origins of Intelligence*. Oxford University Press, Oxford.
Byrne, R.W. (2002). The primate origins of human intelligence. In R.J. Sternberg and J.C. Kaufman (eds), *The Evolution of Intelligence*. Erlbaum, Mahwah, NJ, pp. 79–96.
Byrne, R. and Whiten, A. (eds) (1988). *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Clarendon, Oxford.
Cairns, R.B. (1979). *Social Development: the Origins and Plasticity of Interchanges*. W.H. Freeman, San Francisco, CA.
Call, J. (1999). Levels of imitation and cognitive mechanisms in orangutans. In S.T. Parker, R.W. Mitchell and H.L. Miles (eds), *The Mentalities of Gorillas and Orangutans*. Cambridge University Press, Cambridge, pp. 316–341.
Call, J. and Tomasello, M. (1994). Production and comprehension of referential pointing by orangutans (*Pongo pygmaeus*). *J. Comp. Psychol.* **108**, 315–329.
Call, J. and Tomasello, M. (1996). The effects of humans on the cognitive development of apes. In A.E. Russon, K.A. Bard and S.T. Parker (eds), *Reaching into Thought: The Minds of the Great Apes*. Cambridge University Press, New York, pp. 371–403.

Call, J. and Tomasello, M. (1999). A non-verbal false belief task: the performance of children and great apes. *Child Dev*. **70**, 381–395.
Carpenter, M., Akhtar, N. and Tomasello, M. (1998). 14- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behav. Dev*. **21**, 315–330.
Carpenter, M., Tomasello, M. and Savage-Rumbaugh, E.S. (1995). Joint attention and imitative learning in children, chimpanzees, and enculturated chimpanzees. *Soc. Dev*. **4**, 217–237.
Calvin, W.H. (2002). Pumping up intelligence: abrupt climate jumps and the evolution of higher intellectual functions during the ice ages. In R.J. Sternberg and J.C. Kaufman (eds), *The Evolution of Intelligence*. Erlbaum, Mahwah, NJ, pp. 87–115.
Carlson, S.M. and Moses, L.J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Dev*. **72**, 1032–1053.
Carlson, S.M., Moses, L.J. and Nix, H.R. (1998). The role of inhibitory processes in young children's difficulties with deception and false belief. *Child Dev*. **69**, 672–691.
Chandler, M. and Lalonde, C. (1996). To an interpretive theory of mind: 5- to 7-year-olds' changing conceptions of mental life. In A.J. Sameroff and M.M. Haith (eds), *The Five to Seven Year Shift: The Age of Reason and Responsibility*. The John D. and Catherine T. MacArthur Foundation Series on Mental Heath and Development. University of Chicago Press, Chicago, pp. 111–139.
Clarke, A.D.B. and Clarke, A.M. (1976). Formerly isolated children. In A.M. Clarke and A.D.B. Clarke (eds), *Early Experience: Myth and Evidence*. Open Books, London.
Clark, E.A. and Hanisee, J. (1982). Intellectual and adaptive performance of Asian children in adoptive American settings. *Dev. Psychol*. **18**, 595–599.
Corballis, M.C. (2002). Evolution of the generative mind. In R.J. Sternberg and J.C. Kaufman (eds), *The Evolution of Intelligence*. Erlbaum, Mahwah, NJ, pp. 117–144.
Cosmides, L. and Tooby, J. (1992). Cognitive adaptations for social exchange. In J.H. Barkow, L. Cosmides and J. Tooby (eds), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford University Press, New York, pp. 163–228.
Cosmides, L. and Tooby, J. (2002). Unraveling the enigma of human intelligence: evolutionary psychology and the multimodular mind. In R.J. Sternberg and J.C. Kaufman (eds), *The Evolution of Intelligence*. Erlbaum, Mahwah, NJ, pp. 145–198.
Cummins, D.D. (1996a). Evidence of deontic reasoning in 3- and 4-year-old children. *Memory Cogn*. **24**, 823–829.
Cummins, D.D. (1996b). Evidence for the innateness of deontic reasoning. *Mind Language*, **11**, 160–190.
Cummins, D.D. (1998a). Social norms and other minds: the evolutionary roots of higher cognition. In D.D. Cummins and C. Allen (eds), *The Evolution of Mind*. Oxford University Press, New York, pp. 28–50.
Cummins, D.D. (1998b). Cheater detection is modified by social rank: the impact of dominance on the evolution of cognitive functions. *Evol. Hum. Behav*. **20**, 229–248.
Custance, D.M., Whiten, A. and Bard, K.A. (1995). Can young chimpanzees imitate arbitrary actions? Hayes and Hayes (1952) revisited. *Behavior*, **132**, 839–858.
Cutting, A.L. and Dunn, J. (1999). Theory of mind, emotion understanding, language, and family background: individual differences and interrelations. *Child Dev*. **70**, 853–865.
Davis, H. (1996). Is there a comapative psychology of implicit mathematical knowledge? *Behav. Brain Sci*. **19**, 250.

Deacon, T.W. (1997). *The Symbolic Species: the Co-evolution of Language and the Brain*. Norton, New York.
Dempster, F.N. (1992). The rise and fall of the inhibitory mechanism: toward a unified theory of cognitive development and aging. *Dev. Rev.* **12**, 45–75.
Denenberg, V.H. and Rosenberg, K.M. (1967). Non-genetic transmission of information. *Nature* **216**, 549–550.
Dennis, W. (1973). *Children of the Crèche*. Appleton-Century Crofts, New York.
de Waal, F.B.M. (1997). *Bonobo: the Forgotten Ape*. University of California Press, Berkeley, CA.
de Waal, F.B.M. (ed.) (2001). *Tree of Origins: What Primate Behavior Can Tell Us about Human Social Evolution*. Harvard University Press, Cambridge, MA.
Dunbar, R.I.M. (1992). Neocortex size as a constraint on group size in primates. *J. Hum. Evol.* **20**, 469–493.
Dunbar, R.I.M. (1995). Neocortex size and group size in primates: a test of the hypothesis. *J. Hum. Evol.* **28**, 287–296.
Dunbar, R.I.M. (2001). Brains on two legs: group size and the evolution of intelligence. In F.B.M. de Waal (ed.), *Tree of Origins: What Primate Behavior Can Tell Us About Human Social Evolution*. Harvard University Press, Cambridge, MA, pp. 173–191.
Eccles, J.C. (1989). *Evolution of the Brain: Creation of the Self*. Routledge, New York.
Finlay, B.L. and Darlington, R.D. (1995). Linked regularities in the development and evolution of mammalian brains. *Science*, **268**, 1579–1584.
Finlay, B.L., Darlington, R.B. and Nicastro, N. (2001). Developmental structure in brain evolution. *Behav. Brain Sci.* **24**, 263–308.
Flavell, J.H., Green, F.L. and Flavell, E.R. (1986). Development of knowledge about the appearance/reality distinction. *Monogr. Soc. Res. Child Dev.* **51** (212).
Flavell, J.H. and Miller, P.H. (1998). Social cognition. In D. Kuhn and R.S. Siegler (vol. eds), *Cognition, Perception, and Language* (vol. 2); in W. Damon (gen. ed.), *Handbook of Child Psychology*. Wiley, New York, pp. 851–898.
Fuster, J.M. (1984). The prefrontal cortex and temporal integration. In A. Peters and E.G. Jones (eds), *Cerebral Cortex, vol. 4, Association and Auditory Cortices*. Plenum, New York, pp. 151–177.
Galef, B.G. Jr (1988). Imitation in animals: history, definition, and interpretation of data. In T. Zentall and B.C. Galef Jr. (eds), *Social Learning*. Erlbaum, Hillsdale, NJ, pp. 3–28.
Gallup, G.G. Jr (1979). *Self-recognition in Chimpanzees and Man: a Developmental and Comparative Perspective*. Plenum, New York.
Geary, D.C. (1995). Reflections of evolution and culture in children's cognition: implications for mathematical development and instruction. *Am. Psychol.* **50**, 24–37.
Geary, D.C. (1998). *Male, Female: the Evolution of Human Sex Differences*. American Psychological Association, Washington, DC.
Geary, D.C. and Flinn, M.V. (2001). Evolution of human parental behavior and the human family. *Parenting Sci. Prac.* **1**, 5–61.
Gibson, K.R. (1991). Myelination and behavioral development: a comparative perspective on questions of neoteny, altriciality and intelligence. In K.R. Gibson and A.C. Petersen (eds), *Brain Maturation and Cognitive Development: Comparative and Cross-cultural Perspectives*. Aldine de Gruyter, New York, pp. 29–63.
Goodall, J. (1986). *The Chimpanzees of Gombe*. Belknap, Cambridge, MA.
Gopnik, A. and Astington, J.W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance–reality distinction. *Child Dev.* **59**, 26–37.

Gottlieb, G. (1991). Experiential canalization of behavioral development: results. *Dev. Psychol.* **27**, 35–39.
Gottlieb, G. (1992). *Individual Development and Evolution: the Genesis of Novel Behavior*. Oxford University Press, New York.
Gottlieb, G. (1998). Normally occurring environmental and behavioral influences on gene activity: from central dogma to probabilistic epigenesis. *Psychol. Rev.* **105**, 792–802.
Gottlieb, G. (2000). Environmental and behavioral influences on gene activity. *Curr. Dir. Psychol. Sci.* **9**, 93–102.
Gould, S.J. (1977). *Ontogeny and Phylogeny*. Harvard University Press, Cambridge, MA.
Gould, S.J. (2002). *The Structure of Evolutionary Theory*. Harvard University Press, Cambridge, MA.
Greenfield, P., Maynard, A., Boehm, C. and Schmidtling, E.Y. (2000). Cultural apprenticeship and cultural change: tool learning and imitation in chimpanzees and humans. In S.T. Parker, J. Langer and M.L. McKinney (eds), *Biology, Brains, and Behavior: the Evolution of Human Development*. School of American Research Press, Santa Fe, NM, pp. 237–277.
Greenough, W.T., Black, J.E. and Wallace, C.S. (1987). Experience and brain development. *Child Dev.* **58**, 539–559.
Hare, B., Call, J., Agentta, B. and Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Anim. Behav.* **59**, 771–785.
Hare, B., Call, J. and Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Anim. Behav.* **61**, 139–151.
Harlow, H. (1959). The development of learning in the Rhesus monkey. *Am. Scient.* 459–479.
Harlow, H.F., Dodsworth, R.O. and Harlow, M.K. (1965). Total isolation in monkeys. *Proc. Natl Acad. Sci.* **54**, 90–97.
Harnick, F.S. (1978). The relationship between ability level and task difficulty in producing imitation in infants. *Child Dev.* **53**, 636–642.
Harnishfeger, K.K. (1995). The development of cognitive inhibition: theories, definitions, and research evidence. In F. Dempster and C. Brainerd (eds), *Interference and Inhibition in Cognition*. Academic Press, New York, pp. 175–204.
Harris, P.L. and Nuñez, M. (1996). Understanding of permission rules by preschool children. *Child Dev.* **67**, 1572–1591.
Hayes, K.J. and Hayes, C. (1952). Imitation in a home-reared chimpanzee. *J. Comp. Physiol. Psychol.* **45**, 450–459.
Heyes, C.M. (1998). Theory of mind in non-human primates. *Behavior. Brain Sci.* **21**, 101–148.
Hirata, S. and Morimura, N. (2000). Naïve chimpanzees' (*Pan troglodytes*) observation of experienced conspecifics in a tool-using task. *J. Comp. Psychol.* **114**, 291–296.
Ho, M.-W. (1998). Evolution. In G. Greenberg and M.M. Haraway (eds), *Comparative Psychology: a Handbook*. Garland, New York, pp. 107–119.
Humphrey, N.K. (1976). The social function of intellect. In P.P.G. Bateson and R.Hinde (eds), *Growing Points in Ethology*. Cambridge University Press, Cambridge, pp. 303–317.
International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
Jacobsen, S.W. (1979). Matching behavior in the young infant. *Child Dev.* **50**, 425–430.

Jenkins, J.M. and Astington, J.W. (1996). Cognitive factors and family structure associated with theory of mind development in young children. *Dev. Psychol.* **32**, 70–78.

Jerison, H.J. (1973). *Evolution of the Brain and Intelligence*. Academic Press, New York.

Jerison, H.J. (2002). On theory in comparative psychology. In R.J. Sternberg and J.C. Kaufman (eds), *The Evolution of Intelligence*. Erlbaum, Mahwah, NJ, pp. 251–288.

Joffe, T.H. (1997). Social pressures have selected for an extended juvenile period in primates. *J. Hum. Evol.* **32**, 593–605.

Kaplan, H., Hill, K., Lancaster, J. and Hurtado, A. M. (2000). A theory of human life history evolution: diet, intelligence, and longevity. *Evol. Anthropol.* **9**, 156–185.

Kochanska, G., Murray, K., Jacques, T.Y., Koenig, A.L. and Vandegeest, K.A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Dev.* **67**, 490–507.

Koluchova, J. (1976). The further development of twins after severe and prolonged deprivation: a second report. *J. Child Psychol. Psychiat. Allied Discipl.* **17**, 181–188.

Legerstee, M. (1991). The role of person and object in eliciting early imitation. *J. Exp. Child Psychol.* **51**, 423–433.

Lehman, E.B., McKinley-Pace, M.J., Wilson, J.A., Savsky, M.D. and Woodson, M.E. (1997). Direct and indirect measures of intentional forgetting in children and adults: evidence for retrieval inhibition and reinstatement. *J. Exp. Child Psychol.* **64**, 295–316.

Leslie, A.M. (1987). Pretense and representation: the origins of 'theory of mind'. *Psychol. Rev.* **94**, 412–426.

Lewis, C., Freeman, N.H., Kyriakidou, C., Maridaki-Kassotaki, K. and Berridge, D.M. (1996). Social influence on false belief access: specific sibling influences or general apprenticeship? *Child Dev.* **67**, 2930–2947.

Lewis, M. and Brooks-Gunn, J. (1979). *Social Cognition and the Acquisition of Self*. Plenum, New York.

Lickliter, R. (1996). Structured organisms and structured environments: development systems and the construction of learning capacities. In J. Valsiner and H. Voss (eds), *The Structure of Learning Processes*. Ablex, Norwood, NJ, pp. 86–107.

Lillard, A. S. (2002). Pretend play as Twin Earth: a social-cognitive analysis. *Dev. Rev.* **21**, 495–531.

Lorsbach, T.C., Katz, G.A. and Cupak, A.J. (1998). Developmental differences in the ability to inhibit the initial misinterpretation of garden path passages. *J. Exp. Child Psychol.* **71**, 275–296.

Luria, A.R. (1973). *The Working Brain: an Introduction to Neuropsychology*. Liveright, New York.

McGrew, W.C. (1992). *Chimpanzee Material Culture: Implication for Human Evolution*. Cambridge University Press, Cambridge.

McKinney, M.L. (1998). Cognitive evolution by extending brain development: on recapitulation, progress, and other heresies. In J. Langer and M. Killen (eds), *Piaget, Evolution, and Development*. Erlbaum, Mahwah, NJ, pp. 9–31.

Meltzoff, A.N. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Dev. Psychol.* **31**, 838–850.

Meltzoff, A.N. and Moore, M.K. (1977). Imitation of facial and manual gestures by human neonates. *Science* **198**, 75–78.

Meltzoff, A.N. and Moore, M.K. (1992). Early imitation within a functional framework: the importance of person identity, movement, and development. *Infant Behav. Dev.* **15**, 479–505.

Menzel, E.W. Jr (1974). A group of young chimpanzees in a 1-acre field: leadership and communication. In A.M. Schrier and F. Stollnitz (eds), *Behavior of Non-human Primates*, vol. 5. Academic Press, New York, pp. 83–153.

Miles, H.L. (1994). ME CHANTEK: the development of self-awareness in a signing orangutan. In S.T. Parker, R.W. Mitchell and M.L. Broccia (eds), *Self-awareness in Animals and Humans: Developmental Perspectives*. Cambridge University Press, Cambridge, pp. 254–272.

Mithen, S. (1996). *The Prehistory of the Mind: the Cognitive Origins of Art, Religion and Science*. Thames and Hudson, London.

Moses, L.J., Carlson, S.M. and Sabbagh, M. (2002). On the specificity of the relation between executive function and theory of mind. Paper presented at International Workshop, Young Children's Cognitive Development: Interrelations among Executive Functioning, Working Memory, Verbal Ability, and Theory of Mind. Hirschberg Castle, Eichstätt, Germany.

Nagell, K., Olguin, K. and Tomasello, M. (1993). Processes of social learning in the tool use of chimpanzees (*Pan troglodytes*) and human children (*Homo sapiens*). *J. Comp. Psychol.* **107**, 174–186.

O'Connor, T.G., Rutter, M., Beckett, C., Keaveney, L. and Kreppner, J.M. and the English and Romanian Adoptees Study Team (2000). The effects of global severe privation on cognitive competence: extension and longitudinal follow-up. *Child Dev.* **71**, 376–390.

Oyama, S. (2000). *The Ontogeny of Information: Developmental Systems and Evolution*, 2nd edn. Duke University Press, Durham, NC.

Parker, S.T. and McKinney, M.L. (1999). *Origins of Intelligence: the Evolution of Cognitive Development in Monkeys, Apes, and Humans*. The Johns Hopkins University Press, Baltimore, MD.

Perner, J. (1991). *Understanding the Representational Mind*. MIT Press, Cambridge, MA.

Perner, J., Frith, U., Leslie, A. and Leekam, S. (1989). Exploration of the autistic child's theory of mind: knowledge, belief, and communication. *Child Dev.* **60**, 689–700.

Perner, J. and Lang, B. (2000). Theory of mind and executive function: is there a developmental relationship? In S. Baron-Cohen, H. Tager-Flusberg and D. Cohen (eds), *Understanding Other Minds: Perspectives from Autism and Developmental Cognitive Neuroscience*, 2nd edn. Oxford University Press, Oxford, pp. 150–181.

Perner, J., Ruffman, T. and Leekam, S.R. (1994). Theory of mind is contagious: You catch it from your sibs. *Child Dev.* **67,** 1228–1238.

Perner, J., Stummer, S. and Lang, B. (1999). Executive functions and theory of mind: cognitive complexity or functional dependence? In P.D. Zelazo, J.W. Astington and D.R. Olson (eds), *Developing Theories of Intention: Social Understanding and Self Control*. Erlbaum, Mahwah, NJ, pp. 133–152.

Peskin, J. (1992). Ruse and representations: on children's ability to conceal information. *Dev. Psychol.* **28**, 84–89.

Piaget, J. (1962). *Play, Dreams and Imitation*. Norton, New York.

Potts, R. (1998). Variability selection in hominid evolution. *Evol. Anthropol.* **7**, 81–96.

Povinelli, D.J. (1995). The unduplicated self. In P. Rochette (ed.), *The Self in Infancy: Theory and Research*. Advances in Psychology, 112. Elsevier, Amsterdam, pp. 161–192.

Povinelli, D.J. (1996). Growing up ape. *Monogr. Soc. Res. Child Dev.* **61** (247), 174–189.

Povinelli, D.J. and Bering, J.M. (2002). The mentality of apes revisited. *Curr. Dir. Psychol. Sci.* **11**, 115–119.

Povinelli, D.J., Bering, J. and Giambrone, S. (2000). Toward a science of other minds: escaping the argument by analogy. *Cogn. Sci.* **24**, 509–541.
Povinelli, D.J. and Eddy, T.J. (1996). What young chimpanzees know about seeing. *Monogr. Soc. Res. Child Dev.* **61** (3), Ser. No. 247.
Povinelli, D.J., Landau, K.R. and Perilloux, H.K. (1996). Self-recognition in young children using delayed vs. live feedback: evidence of a developmental asynchrony. *Child Dev.* **67**, 1540–1554.
Povinelli, D.J., Nelson, K.E. and Boysen, S.T. (1992). Comprehension of role reversal in chimpanzees: evidence of empathy. *Anim. Behav.* **43**, 633–640.
Povinelli, D.J. and Simon, B.B. (1998). Young children's understanding of briefly vs. extremely delayed images of the self: emergence of the autobiographical stance. *Dev. Psychol.* **34**, 188–194.
Premack, D. and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* **4**, 515–526.
Provence, S. and Lipton, R.C. (1962). *Infants in Institutions: a Comparison of their Development with Family-reared Infants during the First Year of Life*. International Universities Press, New York.
Raff, R.A. (1996). *The Shape of Life: Genes, Development, and the Evolution of Animal Form*. University of Chicago Press, Chicago.
Reaux, J.E., Theall, L.A. and Povinelli, D.J. (1999). A longitudinal investigation of chimpanzee's understanding of visual perception. *Child Dev.* **70**, 275–290.
Repacholi, B.M. and Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Dev. Psychol.* **33**, 12–21.
Ressler, R.H. (1966). Inherited environmental influences on the operant behavior of mice. *J. Comp. Physiol. Psychol.* **61**, 264–267.
Rilling, J.K. and Insel, T.R. (1999). The primate neocortex in comparative perspective using magnetic resonance imaging. *J. Hum. Evol.* **37**, 191–223.
Ruffman, T., Perner, J., Naito, M., Parkin, L. and Clements, W.A. (1998). Older (but not younger) siblings facilitate false belief understanding. *Dev. Psychol.* **34**, 161–174.
Russell, J., Mauthner, N., Sharpe, S. and Tidswell, T. (1991). The 'windows tasks' as a measure of strategic deception in preschoolers and autistic subjects. *Br. J. Dev. Psychol.* **9**, 331–349.
Sanders, R.J. (1985). Teaching apes to ape language: explaining the imitative and non-imitative signing of a chimpanzee (*Pan troglodytes*). *J. Comp. Psychol.* **99**, 197–210.
Sibulkin, A.E. and Uzgiris, I.C. (1978). Imitation by preschoolers in a problem-solving situation. *J. Genet. Psychol.* **132**, 267–275.
Skeels, H.M. (1966). Adult status of children with contrasting early life experiences. *Monogr. Soc. Res. Child Dev.* **31** (3, Ser. No. 105).
Smendeferi, K., Damasio, H., Frank, R. and Van Hoesen, G.W. (1997). The evolution of the frontal lobes: a volumetric analysis based on three-dimensional reconstructions of magnetic resonance scans of human and ape brains. *J. Hum. Evol.* **32**, 375–388.
Smith, P.K. (1998). *The Theory of Mind Acquisition Support System: Social Origins of Theory of Mind*. Paper presented at Hang Seng Conference on Evolution of Mind, Sheffield, UK.
Spence, K.W. (1937). Experimental studies of learning and higher mental processes in infra-human primates. *Psychol. Bull.* **34**, 806–850.
Spitz, R. (1945). Hospitalism: an inquiry into the genesis of psychiatric conditions in early childhood. *Psychoanal. Study Child* **1**, 53–74.
Stenhouse, D. (1974). *The Evolution of Intelligence*. Barnes and Noble, New York.

Stanford, C.B. (1998). The social behavior of chimpanzees and bonobos: empirical evidence and shifting assumptions. *Curr. Anthropol.* **39**, 399–420.
Stevenson, J.C. and Williams, D.C. (2000). Parental investment, self-control, and sex differences in the expression of ADHD. *Hum. Nature* **11**, 405–422.
Suddendorf, T. and Whiten, A. (2001). Mental evolution and development: evidence for secondary representation in children, great apes and other animals. *Psychol. Bull.* **127**, 629–650.
Suomi, S. and Harlow, H. (1972). Social rehabilitation of isolate-reared monkeys. *Dev. Psychol.* **6**, 487–496.
Thorpe, W. (1956). *Learning and Instincts in Animals*. Methuen, London.
Tobias, P.V. (1987). The brain of *Homo habilis*: a new level of organization in cerebral evolution. *J. Hum. Evol.* **16**, 741–761.
Tomasello, M. (1990). Cultural transmission in the tool use and communicatory signaling of chimpanzees? In S.T. Parker and K.R. Gibson (eds), *'Language' and Intelligence in Monkeys and Apes*. Cambridge University Press, Cambridge, pp. 274–311.
Tomasello, M. (1996). Do apes ape? In C. Heyes and B. Galef (eds), *Social Learning in Animals: the Role of Culture*. Academic Press, San Diego, pp. 319–346.
Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Harvard University Press, Cambridge, MA.
Tomasello, M. (2000). Culture and cognitive development. *Curr. Dir. Psychol. Sci.* **9**, 37–40.
Tomasello, M. and Call, J. (1997). *Primate Cognition*. Oxford University Press, New York.
Tomasello, M., Davis-Dasilva, M., Camak, L. and Bard, K.A. (1987). Observational learning of tool use by young chimpanzees. *J. Hum. Evol.* **2**, 175–183.
Tomasello, M., Kruger, A.C. and Ratner, H.H. (1993). Cultural learning. *Behav. Brain Sci.* **16**, 495–511.
Tomasello, M., Savage-Rumbaugh, S. and Kruger, A.C. (1993). Imitative learning of actions on objects by children, chimpanzees, and enculturated chimpanzees. *Child Dev.* **64**, 1688–1705.
Tooby, J. and Cosmides, L. (1992). The psychological foundations of culture. In J.H. Barkow, L. Cosmides and J. Tooby (eds), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford University Press, New York, pp. 19–139.
Trivers, R. (1972). Parental investment and sexual selection. In B. Campbell (ed.), *Sexual Selection and the Descent of Man*. Aldine de Gruyter, New York, pp. 136–179.
Uzgiris, I.C. and Hunt, J. McV. (1975). *Assessment in Infancy: Ordinal Scales of Psychological Development*. University of Illinois Press, Urbana, IL.
Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W. *et al.* (2001). The sequence of the human genome. *Science*, **291**, 1304–1351.
Vinter, A. (1986). The role of movement in eliciting early imitations. *Child Dev.* **57**, 66–71.
Want, S. C. and Harris, P.L. (2001). Learning from other peoples' mistakes: causal understanding in learning to use a tool. *Child Dev.* **72**, 431–443.
Want, S. C. and Harris, P.L. (in press). How do children ape? Applying concepts from the study of non-human primates to the developmental study of 'imitation' in children. *Dev. Sci.*
Wason, P. (1966). Reasoning. In B.M. Foss (ed.), *New Horizons in Psychology*. Penguin, Harmondsworth, pp. 135–151.
Wellman, H.M. (1990). *The Child's Theory of Mind.* MIT Press, Cambridge, MA.

Whiten, A. (1996). Imitation, pretense, and mindreading: secondary representation in primatology and developmental psychology? In A.E. Russon, K.A. Bard and S.T. Parker (eds), *Reaching into Thought: the Minds of the Great Apes*. Cambridge University Press, New York, pp. 300–324.

Whiten, A. (1998). Imitation of sequential structure of actions by chimpanzees (*Pan troglodytes*). *J. Comp. Psychol.* **112**, 270–281.

Whiten, A. and Byrne, R.W. (1988). The manipulation of attention in primate tactical deception. In R.W. Byrne and A. Whiten (eds), *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Clarendon Press, Oxford, pp. 211–223.

Whiten, A. Custance, D.M., Gömez, J.C., Teixidor, P. and Bard, K.A. (1996). Imitative learning of artificial fruit processing in children (*Homo sapiens*) and chimpanzees (*Pan troglodytes*). *J. Comp. Psychol.* **110**, 3–14.

Whiten, A., Goodall, J., McGrew, W.C., Nishida, T. *et al.* (1999). Cultures in chimpanzees. *Nature* **399**, 682–685.

Whiten, A. and Ham, R. (1992). On the nature and evolution of imitation in the animal kingdom: reappraisal of a century of research. In P. Slater, J. Rosenblatt, C. Beer and M. Milinski (eds), *Advances in the Study of Behavior*, vol. 21. Academic Press, NewYork, pp. 239–283.

Wilson, S.P. and Kipp, K. (1998). The development of efficient inhibition: evidence from directed-forgetting tasks. *Dev. Rev.* **18**, 86–123.

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

Wrangham, R.W. (2001). Out of the *Pan*, into the fire: how our ancestors' evolution depended on what they ate. In F.B.M. de Waal (ed.), *Tree of Origins: What Primate Behavior Can Tell Us About Human Social Evolution*. Harvard University Press, Cambridge, MA, pp. 119–143.

Wyles, J.S., Kunkel, J.G. and Wilson, A.C. (1983). Birds, behavior and anatomical evolution. *Proc. Natl Acad. Sci. USA* **80**, 4394–4397.

Wynn, T. (in press). Archaeology and cognitive evolution. *Behav. Brain Sci.*

Yakovlev, P.I. and Lecours, A.R. (1967). The myelenogenetic cycles of regional maturation of the brain. In A. Minkowski (ed.), *Regional Development of the Brain in Early Life*. Blackwell, Oxford, pp. 3–70.

Youngblood, L.M. and Dunn, J. (1995). Individual differences in young children's pretend play with mother and sibling: links to relationships and understanding of other people's feelings and beliefs. *Child Dev.* **66**, 1472–1492.

Zelazo, P.D. and Frye, D. (1997). Cognitive complexity and control: a theory of the development of deliberate reasoning and intentional action. In M. Stamenov (ed.), *Language Structure, Discourse, and the Access to Consciousness*. John Benjamins, Amsterdam, pp. 113–153.

8

Where Is 'The Other' in the Self? Multiplicity, Unity and Transformation of the Self from a Developmental Standpoint

INGRID E. JOSEPHS[1] AND HEDDA RIBBERT[2]

[1]Department of Clinical Psychology and Personality, University of Nijmegen, The Netherlands, and [2]Centre for Psychiatry and Psychotherapy, University of Bochum, Germany

In psychological research, we find a basic distinction between an 'other-exclusive' and an 'other-inclusive' concept of the self. Whereas the first conceptualisation separates self and other and thus defines the self in a self-contained way, the second conceptualisation regards the relation to the other as constitutive for the self: I am who I am and I know who I am through my relation to the social world. The latter thinking is, however, not new. It has, for instance, roots in George Herbert Mead's (e.g. 1934) theorising and in the dialogical philosophy of Martin Buber (e.g. 1923), to mention only two scholars here.

However, social constructionism and other postmodern 'psychologies' have given rise to the assumption that the other is far more important for individual psychological functioning — including and foremost the self — than psychology has ever thought. In its most extreme version, the social other is celebrated as an all-encompassing explanatory principle, whereas the individual is completely given up or forcefully fought against as a unit of analysis.

In contrast to this often ideologically fortified standpoint, we want to outline a more integrated picture here, not only relating older and newer approaches

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

but also building a bridge between perspectives which have developed in splendid isolation from each other, due to their affiliation to separate subdisciplines within psychology. Furthermore, we start from a developmental perspective, as we are convinced that a phenomenon can only be properly understood in its process of becoming, in its active maintenance and change, which are both outcomes of development.

BASIC ASSUMPTIONS: SOCIAL GUIDANCE AND PERSONAL CONSTRUCTION OF THE SELF

Our most basic theoretical standpoint is that the development of the self is *socially guided* and *personally constructed*. In other terms, development of the self is at the same time a construction process by the person *and* it is directed/constrained (vs. determined) by the social world, i.e. by social others. In still other terms, the person actively constructs his/her own quasi-autonomous personal world (here the self) out of the contact with the guiding and directing (i.e. constraining) social world. This basic claim is crucial, as it does *not* allow for any deterministic, linear assumptions of the following kind: (a) the social world creates (makes) the self (as some social constructionists and behaviourists, too, would argue); or (b) the social world has an (although quite miraculous) 'effect' or 'influence' on the self (as the mainstream of psychology usually argues in its variables-orientated approach).

In elaborating the first part of this claim, i.e. social guidance, we will refer to a prominent sociogenetic theory of ontogenetic human development (including self-development) over the life course, which is at the same time considered to be a middle-level evolutionary theory, namely *attachment theory*, as founded by John Bowlby (1969, 1973) and elaborated by many others up to now. One cornerstone of attachment theory is the assumption of *internal working models of self and other* emerging from social interactions. The basic questions here are: How do I come to know who I am from the interaction with others?; and, How are potentially heterogeneous social interactions on the *interpersonal plane* transformed into and operating on the *intramental plane*? We will see that these central questions are asked not only by Bowlby and others within the realm of attachment theory, but also by famous sociogenetic thinkers (in particular, George Herbert Mead), whose work Bowlby did not know (but could have known). The process of 'translation' of the intermental to the intramental plane, however, cannot be explained without taking the active, agentive and constructive role of the person seriously (the second part of our basic claim) — an aspect not fully covered by attachment theory, as we will see. The opposite solution, namely to conceptualise the person in a passive way and, accordingly, a person's self as an accurate mirror of various relationships,

has to be ruled out from the very beginning, because coordinated psychological functioning can hardly be explained from such a perspective.

Furthermore, attachment theory has yet failed to give an answer to the question of how multiple (including opposite) internal working models of self and others, resulting from multiple and potentially heterogeneous interactions, are internally related and how they operate in real life in a non-chaotic way. In order to give a first and tentative answer, we will build a bridge to the theory of the *dialogical self* (e.g. Hermans and Kempen 1993; Hermans, 1996b), a theory which developed independently of attachment theory within a different subdiscipline of psychology (personality research). How such an answer could be given will be illustrated with the help of an example from the context of psychotherapy, the latter being considered as an important area of self-development.

THE SELF FROM THE PERSPECTIVE OF ATTACHMENT THEORY

BASIC CLAIMS OF ATTACHMENT THEORY

The concept of attachment has found its solid place in contemporary developmental psychology. It is one of the very few obligatory terms that all US undergraduates need to know in order to be considered educated in psychology. The originator of attachment theory was the English psychoanalyst John Bowlby (1907–1990). The theory was created by him over his life course, especially from the late 1930s to the late 1970s. Although starting as a psychoanalyst, Bowlby was interested in considering and integrating many theoretical frameworks throughout his life: systems theory, evolutionary thinking, ethological methodology, Piaget's account of cognitive development — all these theoretical inputs kept his theory in a process of steady development.

Attachment theory emphasises the central role of relationships in human development over the life-course (Bowlby 1969/1982, 1973, 1980). Beginning in infancy and continuing throughout the life-course, an individual's mental health is seen as intimately tied to relationships with attachment figures who give emotional support and protection. Human attachment relationships, according to Bowlby (1979), are regulated by a behavioural-motivational system that develops in infancy. This system monitors physical proximity and psychological availability of a 'stronger and wiser' attachment figure, usually the mother, and activates and regulates attachment behaviour directed towards that figure. As long as an attached individual feels at ease, the attachment figure functions as a secure base of operations whose supportive presence fosters exploration, play or other social behaviours. When the attached

individual feels afraid, however, exploratory goals are overridden by the need to seek refuge with, and reassurance from, the attachment figure, especially if the attached individual is an infant or young child. An individual's attachment to one or a few specific figures therefore becomes most visible under conditions of perceived threat. By seeking an attachment figure's protection, immature offspring are believed to increase the likelihood of their survival and reproductive success (note here that modern evolutionary thinking diverges from this — according to their view — simplistic notion, in which individual well-being and psychological health are the central issues).

STRENGTH AND PROBLEMS OF THE CONCEPT OF THE INTERNAL WORKING MODEL

Beyond infancy, individuals are supposed to construct *internal working models* from the experienced interaction patterns with their principal attachment figures. These internal working models are conceived as 'operable' models of self and attachment partners, based on their joint relationship history. Thus, these models are *representations* (Who am I in the relation to X? Who is X in the relation to me?), but they also *promote further processes*. They serve to regulate, interpret and constrain behaviour, thoughts and feelings. A working model of self as valued and competent is constructed in the context of parents (theoretically and empirically mostly restricted to mothers) as emotionally available, but also as supportive of exploratory activities. In contrast, a working model of self as devalued and incompetent is the counterpart of a working model of parents as rejecting or ignoring of attachment behaviour and/or interfering with exploration. In sum, the basic theorising behind these assumptions is that concrete relationship experiences (belonging to a dyad) are somehow translated into intramentally operating models of the self. However, the way in which this translation is phrased is a crucial point.

Internal Working Model as Quality of the Individual or Quality of a Relation?

The internal working model representing a dynamic relationship can be immediately turned into a construct of categorical property that is assumed to belong to a person and determines his/her conduct. If the working model is assumed to be a property of the person (a trait, rather than a relationship between persons), it becomes similar to any other personality characteristic. Psychology has usually transformed concepts that originally refer to relationships, to depict assumed static properties of the individual. This is often accomplished by typologies — a particular generally labelled property becomes viewed as occurring in different classes (categories).

Attachment theory has proceeded in this way. The translation of a dynamic relationship into a static property usually leads to the attribution of

causal properties to the assumed static entity. Thus, a static quality that was abstracted out of the complex process is turned into a causal agent, which is seen to 'predict' some future outcome (note that these assumed predictive qualities and the 'operative' characteristics of the working model mentioned above are not the same). In attachment theory, Mary Ainsworth's (and attachment researchers' procedure until today) well-known interpretation of *characteristics of dyadic interaction* between a mother and her child as a (trait-like) *characteristic of the child* is an example for this procedure. In her '*strange situation*' experiment (e.g. Ainsworth, Bell and Stayton 1971), in which she observed the reunion pattern between mothers and very young children after a situation of stressful separation in the laboratory, children (and their respective working models) have become classified as A (avoidant), B (secure) or C (ambivalent), according to their behaviour in the reunion phase. Later a fourth category was created (D, disorganised), and since then attachment theory has operated with the four versions of internal working models (solely attributed to the person vs. dyad) throughout the life-span. Different verbal assessments of internal working models in preschoolers, children, adolescents and adults basically come to the same classification of internal working models.

Problems of Change and Multiplicity

This conceptualisation has to face at least two problems. The first problem is the question of developmental change of working models. Note here that stability of the internal working model over the life-course is in some sense the default assumption in attachment theory, despite the rootedness of internal working models in relationships and relationship experiences, which can be assumed to vary. The second problem is the question of how multiple (including discordant) working models resulting from multiple (including opposite) relationships can be theoretically conceived. These questions (change and multiplicity) are discussed separately in the literature, with many more empirical (vs. theoretical) answers given to the first than to the second question.

A link between the two questions could be conceived precisely in addressing the question of how the dynamic interplay of a supposed multiplicity of internal working models leads to the creation of novelty, in other terms to developmental transformation of working models, including the construction of novel working models.

THE PERSON AS PASSIVE OWNER OF A SELF

In attachment theory's attempts to answer both questions — change and multiplicity — up to now, the person as such (in other terms, the 'owner' of the respective working model) is conceptualised in a rather passive way.

Change of working models is basically assumed to result from new social environmental conditions — for better or worse. Yet to do justice to Bowlby, he himself considered the therapeutic talking-through of the working-model (which is certainly not a passive activity) in a safe and unthreatening interpersonal environment as a condition for change, a conceptualisation which shares notable, yet still unexplored, similarities to Carl Rogers's reasoning. One problem of attachment theory in general in recent years might have been that so much attention has been directed to empirical research, in order to find age-adequate methods to assess the internal working model longitudinally over the life course, that theoretical speculations or even elaborations and innovations remained in the background of attention. With regard to the conceptual problem of how an assumed multiplicity of internal working models might be related, or even integrated, there is even less clarity. Thus, the question of how both unity and diversity can be constructively conceptualised is a big puzzle for attachment researchers. Throughout the 1000 pages of the recent *Handbook of Attachment* (Cassidy and Shaver 1999), it becomes clear that this issue belongs to the most crucial topics for future research. In summary, the following tentative solutions are offered in the literature:

1. The person is assumed to select from multiple working models — depending on situational requirements (component model). The relational qualities of the singular working models are retained in this conceptualisation, although no integration (e.g. in the form of a hierarchy) is conceptualised.
2. Attachment classifications are not regarded as mutually exclusive categories, but as dimensions. Thus, a person can have a B classification, while sharing A and C 'portions' to a lesser extent. This mix-model (average model), however, is theoretically hardly viable, as it detaches the working model from its relationship quality in an extreme form.
3. Some sort of hierarchy is assumed without clarifying where this hierarchy is coming from and how it is operating. This model can take two forms: in the first version, generalised relationship experiences resulting in a generalised working model are assumed to operate at the top (relationship qualities are not retained) and specific working models (father, spouse) at the bottom. In the second version, the working model resulting from the relation to the mother is considered to be at the top, with other internal working models at the bottom.

All models are not convincing at either a theoretical level or at the level of psychological functioning. Besides other problems, they are all static and do not offer any theoretical potential to explain developmental change and transformation in a straightforward way.

IN SEARCH OF AN ANSWER TO THE ISSUE OF MULTIPLICITY: MEAD AND THE 'GENERALISED OTHER'

In order to elaborate the question of multiplicity and change, we want to turn to George Herbert Mead's (1934) notion about the social nature of the self, although Bowlby was apparently not aware of his work. Central to Mead's view, which we can only refer to in a hypersimplistic version here, is that a self can only develop when the individual adopts the stance of 'the other' toward him/herself. Mead (1913, 1934) claimed that young children learn about themselves from the responses of important others to their social acts: 'Thus the child can think about his conduct as good or bad only as he reacts to his own acts in the remembered words of his parents' (Mead 1913, p. 377).

With 'the other', Mead refers not only to particular significant others, e.g. the parents, but also to social groups and social role and, as most abstracted, to the community in general — leading to what he called '*the generalised other*'.

> I have pointed out, then, that there are two general stages in the full development of the self. At the first of these stages, the individual's self is constituted simply by an organization of the *particular attitudes of other individuals* toward himself and toward one another in the specific social acts in which he participates with them. But at the second stage in the full development of the individual's self that self is constituted not only by an organization of these particular attitudes, but also *by an organization of the social attitudes of the generalized other or the social group as a whole* to which he belongs (Mead 1934, p. 158).

Thus, in line with some attachment researchers, Mead assumes some general (or generalised) self as an abstraction from concrete relation-experiences. This is what is called ME in the total I–ME dynamic which makes up Mead's 'self'. The ME is not meant to be stable, yet it is clearly the conventional part of the self, and thus significant changes are anticipated at a rather slow tempo. Self-innovation, which leads to a restructuring of the ME, is attributed to the I.

> The 'I' is the response of the individual to the attitude of the community as this appears in his own experience. His response to that organized attitude in turn changes it. As we have pointed out, this is a change which is not present in his own experience until after it takes place. The 'I' appears in our experience in memory. It is only after we have acted that we know what we have done; it is only after we have spoken that we know what we have said (Mead 1934, p. 196).

Thus, the ME is not only transformed in interaction with the social environment — through assuming social roles and carrying out actions appropriate within such roles — but also through the supposed intramental activity of an I–ME feedback loop, in which the activity of the I can be never captured while it is happening, but only reconstructed afterwards through observing the changes in the ME. Whereas the first cycle covers

conventionalisation, the second covers innovation. Yet both cycles co-constrain each other.

For our purposes, Mead's theory is productive and unproductive at the same time. It is productive in the way that it gives a more thoughtful description of the establishment and change of the self, assuming intramental and intermental activity at the same time. It also tackles the problem of stability and change by offering the interplay of the I–ME theoretical dyad. It is not productive in assuming a generalised self, in which particular relationship-experiences are fused: there is a lack of assuming particular others next to the generalised other. Thus, the problem of unity and diversity of the self is solved too easily.

THE DIALOGICAL SELF: MULTIPLICITY, UNITY AND DEVELOPMENTAL TRANSFORMATION

The theory of the *dialogical self* (e.g. Hermans and Kempen, 1993; Hermans, 1996b) interprets the central problem for attachment theory as a central given, and as a point of departure — multiplicity and change. The self can be seen as multitude of *I positions* that have their origin in social relationships. The I is not one, but many. From my position of 'I as a partner', I have a different story to tell than from my position of 'I as a psychologist' or 'I as a mother'. The values I attribute to my life, the topics I select as meaningful, my feeling and thinking in general: they all can potentially differ, depending on my respective I-positions in the here-and-now.

Yet the dialogical self substantially differs from fashionable assumptions of a fragmented or patchwork self, which can be described in terms of what we called a 'component model' before. The principle of *dialogue* is an elaboration and specification of (otherwise fragmented) multiplicity: first, a relation is assumed between the components, and second, this relation is a dynamic one and can potentially lead to the transformation of the whole. The I-positions:

> . . . are organized in an imaginal landscape. In this conception, the I has the possibility to move, as in space, from one position to the other in accordance with changes in situation and time. The I fluctuates among *different, and even opposed, positions* and has the capacity to imaginatively endow each position with a voice so that *dialogical relations* between characters in a story, involved in a process of *question and answer, agreement and disagreement* can be established. Each character has a story to tell about its own experiences from its own stance. These characters exchange information about their respective MEs, resulting in a complex, narratively-structured self. In this multiplicity of positions, *some positions may become more dominant than others*, so that the voices of the less dominant positions may be subdued (Hermans 1996a, pp. 10–11, added emphases).

Thus, each I-position creates a 'voice' which relates to other voices (of other I-positions) in a dynamic relation of dialogicality. Opposition of hierarchically different voices (dominance–subdominance relation), together with the dialogical interaction between those voices, is a condition under which developmental transformation (e.g. dominance reversal) is supposed to happen.

Usually, the notion of opposites and oppositionality is a burden for psychological theorising, at least since the decades under which psychology has operated under a methodological dogma. To describe oneself at the same time as A and non-A is — methods-wise — impossible for psychology to capture (think of questionnaires here and other methods that follow the bivalent logic). In the dialogical self, however, oppositionality is one condition for change. For the study of the dialogical self, then, conditions under which the processes are accessible need clarification. It is no coincidence that Hermans has built his model of the dialogical self on the empirical evidence from the psychotherapy process. Here the dialogical self can be effectuated by asking clients to describe different sides of their own personality and inviting them to formulate from each side a separate meaning system in their own terms. It is one goal of psychotherapy to initiate a dialogue between these opposite positions. This dialogue is assumed to be a starting point for the transformation of the self-system as such.

DIALOGICAL TRANSFORMATION OF THE MULTIPLE SELF: A CASE FROM PSYCHOTHERAPY

The following example illustrates how change of the system is resulting from a dialogue between opposite components in the course of a psychotherapy process. Hermans and Hermans-Jansen (1995) report the story of Mary, a 33 year-old woman suffering from childhood abuse by her father. Mary married a man whom she loved very much, yet at times felt strong disgust for him. Sometimes, often unexpectedly, she felt a sudden fierce aggression toward him that was entirely beyond her control. There were times when she felt like a witch, an alien experience that frightened her, particularly when the witch took almost total possession of her. At the beginning of the psychotherapy process, the 'usual Mary' and 'Mary as a witch' were operating as two opposite parts of Mary's self-system. Their stories were very different, as were the meanings and emotions attributed to them. Furthermore, the witch had the potential to control Mary from time to time, which implies a dominance relation. At the end of psychotherapy, however, Mary is no longer able to articulate a clear witch-voice. The 'usual Mary' and the 'Mary as a witch' have been transformed in a qualitative manner. The witch has lost its clear-cut characteristics and became functional for the sake of Mary (as a protector for Mary in dangerous situations). A new dominance structure occurred, in which Mary had the

control and the witch became an integrated part of Mary (for a similar example, see also Miltenburg and Singer 1999; for a psychological interpretation, see Valsiner 1999).

In the course of the psychotherapy process, Mary was instructed to listen to the witch and to bring this voice in dialogue with her usual voice. This made it possible for Mary to recognise that the witch-voice was not an enemy but a part of the system that was fighting for legitimate and important goals, yet in a non-acceptable way. Thus, this internal dialogue (e.g. in the form of a diary), and the dialogue between Mary and the psychotherapist, enabled Mary to actively transform her self-system in a productive way.

SOME CONCLUSIONS

The self is social in its origin, yet it is personally constructed and worked upon, basically through semiotic activity (internal processes) which are guided by interpersonal processes (dialogue between client and psychotherapist, here). Whereas the first part of the statement could be illustrated by attachment theory, the second part owes to observations from the theory of the dialogical self. To psychotherapists, the insights of the theory of the dialogical self might not be new. In fact, there are other theories of psychotherapy that start from the assumption of multiplicity of the self, and actively work with this assumption in the process of therapy. Yet *academic psychology* has had (and still has!) a hard time taking into account these dynamic processes of multiplicity, dialogicality and oppositionality, partly because the focus is always on what is empirically do-able within the generally accepted framework of research. Needless to say, issues of multiplicity and its organisation, and the dynamic outcome of oppositionality, are hard to conceive in a convincing way within the methodological credo of strict statistical analyses. In the future we need to to show exactly how the (internal and external) dialogical process works, and under which conditions it can lead to change or likewise block change. For this we need convincing theoretical models, which have still to be constructed.

REFERENCES

Ainsworth, M.D.S., Bell, S.M. and Stayton, D.J. (1971). Individual differences in Strange-Situation behavior of 1 year-olds. In H. R. Schaffer (ed.), *The Origins of Human Social Relations*. Academic Press, New York, pp. 17–52.

Bowlby, J. (1969/1982). *Attachment and Loss, vol. 1. Attachment*. Basic Books, New York.

Bowlby, J. (1973). *Attachment and Loss, vol. 2. Separation*. Basic Books, New York.

Bowlby, J. (1979). *The Making and Breaking of Affectional Bonds*. Tavistock, London.

Bowlby, J. (1980). *Attachment and Loss, vol. 3. Loss*. Basic Books, New York.
Buber, M. (1923/1994). Ich und Du [I and Thou]. In *Das dialogische Prinzip*. Verlag Lambert Schneider, Gerlingen, pp. 7–38.
Cassidy, J. and Shaver, P.R. (eds) (1999). *Handbook of Attachment*. Guilford Press, New York.
Hermans, H.J.M. (1996a). Opposites in a dialogical self: constructs as characters. *J. Constructivist Psychol*. **9**, 1–26.
Hermans, H.J.M. (1996b). Voicing the self: from information processing to dialogical interchange. *Psychol. Bull*. **119**, 31–50.
Hermans, H.J.M. and Hermans-Jansen, E. (1995). *Self-narratives. The Construction of Meaning in Psychotherapy*. Guilford, New York.
Hermans, H.J.M. and Kempen, H.J.G. (1993). *The Dialogical Self: Meaning as Movement*. Academic Press, San Diego, CA.
Mead, G.H. (1913). The social self. *J. Philos*. **10**, 374–380.
Mead, G.H. (1934). *Mind, Self, and Society*. University of Chicago Press, Chicago, IL.
Miltenburg, R. and Singer, E. (1999). Culturally mediated learning and the development of self-regulation by survivors of child abuse: a Vygotskian approach to the support of survivors of child abuse. *Hum. Dev*. **42**, 1–17.
Valsiner, J. (1999). I create you to control me: a glimpse into basic processes of semiotic mediation. *Hum. Dev*. **42**, 26–30.

PART IV

Pathologies of the 'Social Brain'

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

9

The Social Brain in Autism

FRED R. VOLKMAR, AMI KLIN, ROBERT T. SCHULTZ, KATARZYNA CHAWARSKA AND WARREN JONES

Child Study Center, Yale University School of Medicine, New Haven, CT, USA

The autistic syndrome was first described in 1943 by Leo Kanner in a remarkably enduring report of 11 children, who exhibited what Kanner suspected to be a be a congenital lack of social orientation and interest. In contrast to their lack of interest in people, these children with early infantile autism had very high levels of interest in the non-social environment, e.g. while rather indifferent to people, they might become highly agitated in response to what seemed to be trivial changes in the inanimate environment. Kanner strongly emphasised the major importance of the social dysfunction in this syndrome and regarded it as an essential feature of the condition. Furthermore, he highlighted the strong contrast to typically developing children, for whom the social world seemed to have such centrality from the first days of life.

Now, nearly 60 years later, the social disability in autism is a profound one, affecting a person's capacity for understanding other people and their feelings, and for establishing reciprocal relationships. This core social disability both defines these conditions and contributes to disability in other areas, e.g. in communication and behaviour. Despite the wide recognition of the central role of the social deficit as a defining feature of autism (Grossman, Carter and Volkmar 1997), a more precise characterisation and quantification of the social dysfunction required to direct neurobiological research in autism is still lacking (Klin *et al*. 2002a). For example, major advances in the genetics of autism have identified candidate susceptibility loci (IMGSAC 1998), but the relationship of social vulnerabilities to genetic vulnerabilities remains to be

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

specified and to intervening brain mechanisms remains to be clarified. As we have recently summarised (Klin *et al*. 2002b), different lines of research point to the need to refine the characterisation of social dysfunction in autism so as to capture essential elements of sociability, which may be disrupted, to differing degrees, in individuals with prototypical autism as well as with the broader manifestations of this condition. These difficulties are expressed in a range of ways — from disabling social anxiety and behavioural rigidity to difficulties in social interaction and interpretation of non-verbal communication (Grossman, Carter and Volkmar 1997).

In this chapter we consider several perspectives on understanding the social brain in autism. The first perspective emphasises the centrality of the social deficit in autism — its nature, pervasiveness and severity. The second section summarises current theoretical views of this deficit (or deficits). The uses and limitations of these views are highlighted. The final section considers potential neural mechanisms and highlights areas important in achieving an integration of both clinical and theoretical perspectives, to more adequately understand the underlying neural basis of autism. We emphasise emerging methods that may better capture the underlying social deficit in autism by focusing on explicit psychological processes and their potential relationship to specific brain systems/mechanisms.

CLINICAL PERSPECTIVES: THE NATURE OF SOCIAL DYSFUNCTION IN AUTISM

SOCIAL DEFICIT AS A DIAGNOSTIC FEATURE

Kanner's emphasis on the centrality of social deficit in autism as a, if not the, central defining feature has been continuously emphasised in the years since his first description of the condition. For example, in his influential synthesis of Kanner's report with subsequent research, Rutter's (1978) definition included unusual (delayed *and* deviant) social development as an essential feature of the condition. Similarly, early epidemiological studies, such as those by Wing and Gould (1979), included social deficits as a core feature of the condition. The development of explicit, official diagnostic guidelines for autism in DSM-III (APA 1980) continued this tradition and social dysfunction continues to be included in both the international (ICD-10; WHO 1990) and DSM-IV (APA 1994) definitions of autism.

Empirical work supporting the centrality of the social deficit has been provided by various investigators, e.g. Siegel *et al*. (1989) reanalysed data about ratings of DSM-III-R criteria for autism using signal detection analysis and found that research social criteria were found to be the most potent predictors of diagnosis. In both DSM-IV and ICD-10, qualitative impairments

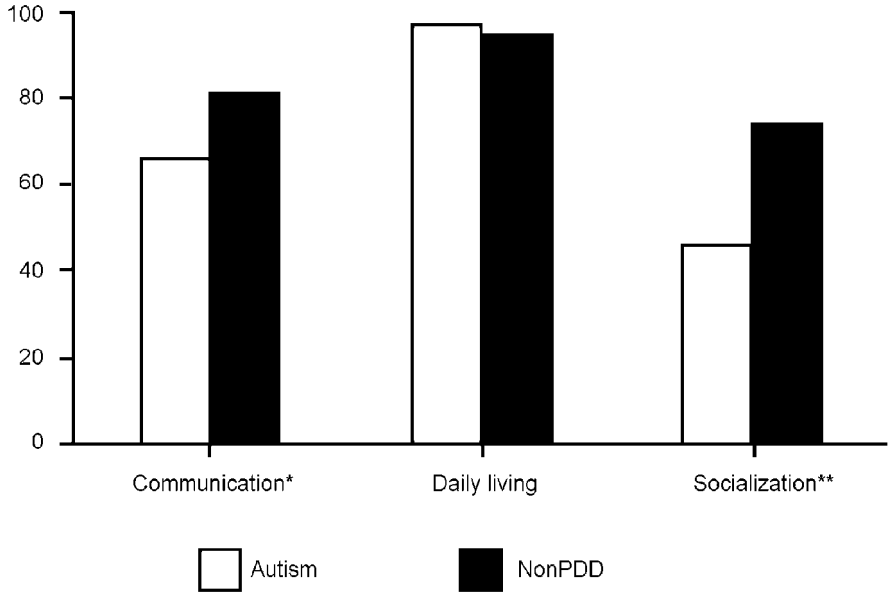


Figure 9.1. Ratios of Vineland Domain scores for Communication, Daily Living and Socialization skills to mental age ($\times 100$) for individuals with autism contrasted to individuals with developmental disorders, $^{*}p < 0.05$, $^{**}p < 0.003$. Data from Volkmar *et al.* (1987)

in social interaction are among the essential diagnostic features; consistent with earlier research, empirical analysis of the field trial data strongly supported the importance and centrality of social deficits in autism and this is embodied in the current definitions by the greater weight placed on social disturbance (Volkmar *et al*. 1994).

In addition to the categorical approaches used for the definition of social deficit, various dimensional approaches have also been employed, e.g. in rating scales and checklists. These rather diverse approaches have included methods based on more normative developmental methods as well as those focused more on deviant social development (Lord 1997). The more dimensional approaches to the definition of social deviance share some general challenges and tensions, e.g. the use of parent report vs. observation, the specification of specific deviant developmental features as opposed to developmental delay. For example, Volkmar and colleagues (1987) utilised a normative measure of adaptive skill and reported that children with autism exhibited much lower than expected social skills in comparison to a mentally handicapped group, even when cognitive level was taken into account (Figure 9.1). Subsequent

work utilised signal detection methods to demonstrate that delays in social skills are, of themselves, the most robust predictors of diagnosis of autism, even when mental age is controlled for (Volkmar *et al.* 1993). Other studies have reported similar results (e.g. Freeman *et al.* 1991; Loveland and Kelley 1991; Rodrigue, Morgan and Geffken 1991).

DEVELOPMENTAL ASPECTS OF SOCIAL DEVELOPMENT IN AUTISM

Early studies of the social development of children with autism (e.g. Ornitz, Guthrie and Farley 1977; Volkmar, Cohen and Paul 1986) were based on parent report rather than direct observation. Subsequent work has confirmed the severity of social deficits in autistic children using observational and other clinical procedures; this work does, however, suggest that there is variability in terms of which social abilities are most impaired (Trad *et al.* 1993; Capps *et al.* 1994; Stone, Ousley and Littleford 1997; Werner *et al.* 2000). In younger children delays or deviance in social development are typical, as children fail to develop reciprocal eye contact and fail to engage in the marked social engagement observed in typically developing infants; in marked contrast, the attention of infants with autism to the *inanimate* environment may be quite exquisite (Stone 1997).

In an attempt to deal with the problems of over-reliance, with parental retrospection on the one hand and the lack of ecological validity of highly artificial experimental procedures on the other, some studies have utilised other methods, e.g. analysis of movies/videos of children as infants. For example, Osterling and Dawson (1994), noted that infants who subsequently exhibited autism were more likely to exhibit fewer social behaviours, less joint attention and more behaviours typical of autism. Prospective studies are rare but also have supported the notion that deficits in social interaction are frequent, even in very young children with autism (Gillberg *et al.* 1990). With the increased interest in genetics and in early diagnosis it is likely that much larger prospective series will be studied in the future (Chawarska, Volkmar and Klin 2002).

Rather typically, the picture of social engagement in very young children with autism is one that has been described as 'aloofness' by Wing (1981), i.e. it is difficult to engage the child in shared routines or social activities. Clearly with increasing developmental level some social skills do emerge, as reflected, for example, in evidence of increasing attachment to parents (Capps *et al.* 1994). Similarly, as children with autism become older they are more differentially responsive to familiar adults and begin to self-monitor (Mundy *et al.* 1986; Volkmar *et al.* 1989; Buitelaar *et al.* 1991, 1992; Shields *et al.* 1996; Mundy and Crowson 1997; Hobson and Lee 1998; Sigman *et al.* 1999). Interestingly, other processes with important social elements seem less likely to be grossly disturbed in autism, e.g. given the apparent congenital difficulties of

autism, one might expect problems in early feeding with resulting failure to thrive but this is, in fact, rather uncommon. In other instances skills with social elements do emerge but in somewhat unusual ways. For example, in typically developing children transitional objects (much-loved blankets, teddy bears, and so forth) often develop in the second year of life and serve important functions for the toddler, e.g. in negotiating transitions. In autism when such objects are observed they differ from more typical ones in two ways — the objects chosen are usually hard rather than soft and it is the category of object rather than the specific object that is important. The autistic child may, for example, carry around a box of cereal and carry it even to bed with him but it is the brand of cereal rather than the box which is important.

The social deficit associated with autism is clearly not a static one (Volkmar *et al.* 1997); changes may occur over both developmental level and chronological age, with the most severe and 'classic' picture observed in younger children. Wing has suggested that, with developmental gain, the social picture changes from one that she characterises as 'aloof' to a more passive stance, i.e. where the child is unlikely to initiate social contact but will respond, and finally, to what she terms an 'active but odd' social style. This latter style is one characterised by one-sided, rather eccentric and poorly coordinated social overtures of the type seen in higher functioning, usually older, individuals; for example, the person may seek new social contact but does it in a highly idiosyncratic and one-sided way. The nature of the processes underlying social gains made over time remains poorly understood. It is unclear whether advances made follow a more predicable or a more atypical developmental sequence (Volkmar, Burack and Cohen 1990; Loveland and Kelley 1991).

Even for the relatively small number of individuals with autism who are able to attain normalcy (typically defined by personal independence and self-sufficiency) social deficits remain and are quite striking (Volkmar and Cohen 1985). Often the highest functioning individuals seem highly motivated to form lasting social relationships, e.g. with members of the opposite sex, but have substantial problems dealing with reciprocal interaction and in generalising the rules of social interchange. As a result, feelings of inadequacy and isolation are rather common (Volkmar *et al.* 1997). Although a growing body of work exists on social skills intervention (National Academy of Sciences 2001) there is, at present, a dearth of studies which validate such treatments empirically.

THEORETICAL APPROACHES TO UNDERSTANDING SOCIAL DYSFUNCTION IN AUTISM

Various approaches have been utilised in the attempt to understand social dysfunction in autism. At the most basic level one approach has been to evaluate and quantify difficulties in social organisation at the level of

symptoms, e.g. as exemplified in DSM-IV (APA 1994; Volkmar *et al*. 1994) or as exemplified in a more dimensional way through the use of instruments like the ADI-R or ADOS-G (see Lord 1997). These approaches are very important in standardising diagnostic approaches but are intrinsically limited in elucidating mechanisms of social dysfunction, i.e. such approaches help us define the area of inquiry but do not provide us with either a normative, developmental context or the opportunity to dissect out highly specific social processes. While symptom-based genetic analyses of some psychiatric disorders, such as obsessive-compulsive disorder, have been productive (Alsobrook *et al*. 1999) it was the potential for relating these to normative tendencies or developmental processes (e.g. the need for symmetry and ordering) that appeared to correspond to a genetically significant subtype of obsessive-compulsive disorder and thus brought these behaviours into a perspective more helpful to development of theoretical models. Similarly, genetic analyses using measures of normative skills whose impairment appears to lead to a syndrome or disability, rather than measures of the disability itself, have been more successful in capturing inheritable vulnerabilities, as in the case of deficits in phonemic awareness in reading disabilities (Grigorenko *et al*. 1997). Despite the intrinsic limitations of the symptom-based approach, emerging work suggests some potential for defining degrees of severity in both affected individuals and family members (Constantino *et al*. 2000).

Another approach has been the search for neurodevelopmental processes that are presumed to represent 'core' (i.e. causative) deficits in autism. Various processes — perceptual, neuropsychological or behavioural — have historically been hypothesised to lie at the root of the social difficulties in autism, e.g. problems in attention (Leekam, Lopez and Moore 2000), perception (O'Neill and Jones 1997) and language (Lord and Paul 1997). While this work has helped to clarify important aspects of behavioural and psychological functioning in autism it has, with the exception of the study of language-communication, not proved particularly helpful in clarifying mechanisms of social disability. Studies of language in autism have repeatedly noted that the degree of language impairment is highly correlated with degree of social dysfunction, although both measures are also highly correlated with overall intellectual levels, which it is totally predictive of social dysfunction, since more able individuals, i.e. the approximately one-third of individuals with IQs in the normal range, have reasonably adequate formal language capacities apart from prosodic and pragmatic skills (Fombonne 1999), and yet the persons have profound social disabilities. Similarly, work on the genetics of the condition has shown that neither language nor IQ deficits aggregate in straightforward ways in family members of probands with autism (Pickles *et al*. 2000), i.e. in a way that would separate these families from families of individuals with other conditions.

Several recent attempts have been made to provide broader theoretical frameworks that might account for the social dysfunction in autism and autism spectrum conditions. Executive functioning (EF) skills include a group of abilities that allow the person to maintain an appropriate problem-solving set in order to attain a goal. Among the various constructs subsumed under EF, planning, and particularly flexibility or set-shifting, are presumed to be the skills most affected in autism (Ozonoff 1997). This hypothesis has great face validity given that individuals with autism are known, for example, to perseverate on inappropriate responses and to have great difficulty in planning and organising their daily affairs. Abnormalities in the brain circuitry subserving EF, particularly the dorsolateral prefrontal cortex, are thought to give rise to the social dysfunction in autism (Pennington and Ozonoff 1996). There are several challenges to this hypothesis, however, including the findings that EF deficits are seen in a number of other disorders (Pennington and Ozonoff, 1996), and such deficits may not correlate well with degree of social disability (Dawson *et al.* 1998). Nevertheless, a small number of studies have already shown differential aggregation of EF deficits in parents (Hughes, Leboyer and Bouvard 1997) and in siblings of autistic probands (Hughes, Plumet and Leboyer 1999).

Another theoretical approach has centred around the construct of central coherence or the capacity to integrate information into coherent or meaningful wholes (Frith 1989). Weak central coherence is presumed to account for characteristic difficulties with attention, appreciation of context and overall meaning. Even though this hypothesis is appealing the hypothesised drive for such central coherence has not been put in either a developmental or neurofunctional context and the supportive literature is sparse and rather limited (Mottron, Peretz and Menard 2000). Indeed, one might just as readily argue that the difficulties in central coherence in autism derive from the underlying social disability, rather than the reverse, e.g. lacking the motivation for social interaction, infants with autism are not drawn to derive meaning from what are usually the organising stimuli and experiences in the first months of life, such as the human face (Vecera and Johnson 1995; Hains and Muir 1996; Caron *et al.* 1997); such experiences presumably serve the typically developing child as important and enduring templates for learning.

Probably the most successful theoretical account for social dysfunction, at least in terms of its research productivity, has been the theory of mind (ToM) hypothesis (Baron-Cohen 1995), which posits that deficits in social interaction arise as a result of the difficulty in conceptualisation of mental phenomena in both self and others, i.e. in the basic foundations of intersubjectivity. Persons with autism are thought to be unable to think about other people's beliefs, intentions, desires and feelings, and thus cannot use this knowledge to explain or predict another person's behaviour (Baron-Cohen 1988). This notion accounts for many aspects of social disability, e.g. the difficulties with implied

meaning and pragmatic language so common in individuals with autism and related disorder. A number of problems arise, however, given the lack of specificity of the hypothesis, the strong relationship of ToM skills to language, and the observation that many higher functioning individuals with autism or Asperger's syndrome are quite socially disabled and yet are able to successfully engage in usual ToM tasks (Bowler 1992; Dahlgren and Trillingsgaard 1996). A further problem arises since it appears that the social deficits in autism are ones that arise, developmentally, before the earliest manifestations of ToM skills (Klin, Volkmar and Sparrow 1992). An important effect of this, and other current theoretical models of autism, has been the attempt in recent years to relate specific social processes to brain mechanisms.

NEURAL MECHANISMS: UNDERSTANDING THE SOCIAL BRAIN IN AUTISM

Despite the limitations of current theoretical perspectives, considerable advances have been made in recent years relative to specific aspects of social dysfunction in autism. For example, a great deal of attention has concentrated on studies of face perception (Klin *et al.* 1999; Schultz *et al.* 2000b). This line of work is quite relevant to autism, given the central role attributed to face perception in the usual process of socialisation. Behavioural work with children with autism has shown deficits in face perception relative to mental-age matched controls (Klin *et al.* 1999). While deficits in face perception are relatively reliable and robust, often the effect size is modest, due to compensatory strategies, e.g. in higher functioning individuals (Boucher and Lewis 1992); nevertheless, even in this group there are functional abnormalities on fMRI (Schultz *et al.* 2000a), with decreased fusiform and increased inferior temporal gyrus activation. Essentially these higher functioning individuals with severe social difficulties treated faces as 'objects'. These findings have highlighted the need to focus not only on results but also on the ways in which individuals with autism perform face perception tasks, and the developmental considerations necessary to interpret performance results. For example, there is some indication that rather than representing a face-specific deficit of presumably neurofunctional basis, the abnormalities in autism may reflect a lack of expertise of autistic individuals with face stimuli (Schultz *et al.* 2000b), which would then probably reflect a lack of repeated engagement with such stimuli early in life due the person's history of social disengagement, or in other words, this functional marker may be more of an 'effect' of autism, as opposed to a cause (Klin, Schultz and Cohen 2000). This would not diminish the utility of measures of face perception as quantifiable indicators of social disability and attempts are being made in this regard (Schultz *et al.* in press). Face perception methods have not yet been used in genetic research in autism.

Nevertheless, there are strong neurofunctional models of face processing skills in typical (Gauthier *et al.* 1999), brain-injured (Adophs *et al.* 1996) and autistic populations (Klin *et al.* 1999; Schultz, Romanski and Tsatsanis 2000a), raising the possibility that a neurofunctional social 'endophenotype' is not too far in the horizon. In this section we focus on selected social processes and their potential neurobiological bases; we emphasise that at present a comprehensive theory of the neurobiology of social dysfunction has not yet appeared — although hopefully the substantial gains made in both more basic work and that focused specifically on autism will make the exposition of such a theory likely in the not so distant future. Before turning to the discussion of specific processes, however, we should briefly review the evidence in favour of a neurobiological aetiology of autism.

Despite the fact that there is now essentially universal agreement on the importance of neural mechanisms in autism, it is striking that for many years after Kanner's original (1943) description there was little or no attention to brain mechanisms. Some of this confusion arose as a result of some aspects of Kanner's original report, e.g. relative to whether autism was a disorder associated with higher levels of parental education. Early attempts to explain autism on the basis of deviant parenting or early experience were discarded only as unequivocal and dramatic evidence of central nervous system involvement emerged, e.g. the observation of high rates of seizure disorders in children with autism (up to 25% of cases), of the persistence of 'primitive' reflexes, and other signs of neurological dysfunction (Minshew, Sweeney and Bauman 1997). Similarly, for many years there was a failure to appreciate the strong role of genetic factors in autism. It became apparent that while autism could be associated with various medical conditions, it was most strongly associated with several genetic conditions and indeed rates of autism in siblings were clearly elevated (Rutter *et al.* 1997).

Various approaches have been used in the study of brain development and functioning in autism. These range from studies of neurochemistry and neuropathology to fMRI and other functional approaches. Some findings, e.g. of elevated peripheral levels of the neurotransmitter serotonin, have been well replicated but their functional significance has remained unclear (Anderson and Hoshiono 1997). Early neuroimaging studies revealed some contradictory findings, e.g. relative to the cerebellum (Courchesne *et al.* 1988), which have not proven readily replicable (Piven *et al.* 1992). Neuropathological studies have suggested abnormalities in various brain regions, particularly limbic area where neurons are more densely packed and have stunted dendritic arborisation and in the cerebellum where there are decreased number of Purkinje and, to a lesser extent, granule cells (Kemper and Bauman 1998).

Several attempts have been made to relate observed behavioural and developmental deficits and/or presumed psychological mechanisms to brain structure and function. For example, one approach has attempted to relate

deficits in executive function to disturbances in the frontal lobes (Ozonoff, Pennington and Rogers 1991); another has focused on aspects of frontal and parietal cortex as well as the neocerebellum (Courchesne 1997). Yet another approach has been concerned with aspects of memory and the limbic system (Bachevalier, 1996; Bachevalier, Malkova and Mishkin 2001). Minshew and colleagues (1997) have also proposed a complex information processing neocortical systems theory. These approaches have variably emphasised some putative basic deficit that is presumed to impact on the individual's various abilities, and all share a concern with aspects of frontal lobe functioning. These diverse theoretical approaches differ in their emphasis on certain processes, e.g. the specific importance of memory as opposed to more general deficit in information processing. Theoretical differences may be less substantial than first appears, since rather different terms are often used to describe rather similar clinical phenomena. However, in at least one case (Bachevalier 1996) the theory has led to the attempt to produce a functional animal model of autism. As with the overarching psychological theories these more neurobiologically-orientated theoretical approaches offer considerable heuristic value. At present, however, empirical work using these approaches is — at least in terms of the social deficit in autism — relatively weak. In the remainder of this chapter, rather than focus on overarching theoretical models, we shall turn our attention to specific social processes and their underlying neurological basis.

SOCIAL AFFECTIVE PROCESSING

In an initial study of ToM abilities, Baron-Cohen *et al*. (1994) utilised single photon emission computed tomography (SPECT) procedures in a task where subjects were asked to imagine mental states as contrasted to non-mental states; they observed increased blood flow in the right orbitofrontal cortex relative to the left frontal poles but did not, unfortunately, report on blood flow in other brain regions. Fletcher *et al*. (1995) observed selective activation of an area of the left medial frontal cortex using a positron emission tomography (PET) approach in which subjects were asked to read stories with/without mental state elements. Happé (1996), in a study of a small number of individuals with Asperger's syndrome, used the same task and found the centre of activation shifted into a region between Brodmann areas 9 and 10. Similar results have been noted by Goel and colleagues (1995); they used a task similar to Fletcher's and found activation in both the left medial frontal lobe and the left temporal lobe.

Our group (Klin, Schultz and Cohen 2000; Schultz, Romanski and Tsatsanis 2000a) has employed a visual procedure to investigate social activation of the brain. In this procedure, subjects observe a series of short animations in which geometric shapes 'interact' and the subject is required to make a judgement of whether the shapes are 'friends' or 'not friends'. This novel procedure has the

advantage of minimising several potential confounding elements, e.g. absence of facial expression and vocalisation revealed activation of the medial prefrontal cortex centred in BA 9, as well as in the amygdala and fusiform gyrus. Similar results, using a PET procedure, were obtained by Castelli and colleagues (2002).

Brothers, Ring and Kling (1990) have proposed that a specialised neural circuit centred on the amygdala, and involving the orbitofrontal cortex, anterior cingulate and superior temporal sulcus, may underlie social cognition. This would be consistent with results noted above as well as with histopathological reports of abnormalities in the amygdala and limbic forebrain, with reduced neuronal size, greater cell packing density and immature-appearing neurons (Kemper and Bauman 1998). In a recent fMRI study, Ring *et al*. (1999) used an embedded figures task and noted difficulties in derivation of relevant information from eye gaze in individuals with Asperger's syndrome/high functioning autism. Significantly elevated activation was observed in the superior temporal gyrus bilaterally, while the pattern of activation in the control group involved greater activation in left amygdala, right insula and left inferior frontal gyrus. The latter are areas of special relevance for social cognition (Brothers, Ring and Kling 1990; Damasio 2001).

These studies are of interest, given the convergence of results with rather different paradigms and methods. Taken together, they suggest selective areas of brain activation in tasks of social judgement or attribution and mentalising in regions including medial prefrontal and maybe orbitofrontal areas, midtemporal structures, and in portions of the amygdaloid complex as well. It must also be noted that these areas are associated with a range of activities, including affective processing (Lane *et al*. 1997a,b) integration of social reasoning, and action (Damasio 1995).

JOINT ATTENTION AND GAZE MONITORING IN AUTISM

Joint Attention

One of the defining criteria of autism in most diagnostic instruments as well as in measures developed for the screening for autism in infancy is deficit of joint attention (JA). On the behavioural level, JA involves episodes in which two people share attention to an object of mutual interest. These episodes may occur when a child follows the attention of an adult by simply 'looking where someone else is looking' (gaze monitoring) or following communicative gestures of others, such as pointing (Guillaume 1971; Murphy and Messer 1977; Butterworth and Jarrett 1991). The child may also initiate such an episode by directing attention of an adult to objects by shifting gaze between the objects and the adult, showing them or pointing to them in order to communicate interest (protodeclarative pointing) (Bates 1979). Initiation and

responding to attention of others, although related, appear to reflect distinct psychological processes (Butterworth 1995; Mundy and Gomes 1998). The ability to respond to others' bids for attention precedes the emergence of the ability to initiate such bids in ontogenesis, each of them predicting a different aspect of language development, and they are poorly intercorrelated (Morisette, Ricard and Decaire 1995; Mundy 1995; Ulvund and Smith 1996; Mundy and Gomes 1998). On the conceptual level, JA behaviours reflect an understanding of other people as intentional, goal-orientated entities (Tomasello 1995; Moore and Corkum 1998; Moore 1999). In this sense, JA constitutes a first step in the ontogeny of social cognition and the development of theory of mind (Mundy *et al.* 1986). Although infants exhibit some rudimentary forms of JA early in the second year of life, these behaviours are at first infrequent, highly context-dependent and their conceptual basis is not fully understood. It is not until about 18–19 months of age that these skills become more robust and functional across contexts (Tomasello 1995; Moore and Corkum 1998).

The ability to share attention with others in early development provides the foundation for communicative and social-cognitive development (Tomasello, Kruger and Ratner 1993; Ulvund and Smith 1996). JA episodes have high functional significance for early language development, including comprehension of language (Mundy and Gomes 1998), production of verbal and non-verbal communicative behaviours (e.g. Baldwin 1995) and novel word learning (Tomasello and Farrar 1986). JA has been also implicated in the phenomenon of social referencing, in which emotional information about an ambiguous object or event is conveyed from adult to infant (Sorce *et al.* 1985).

Gaze Monitoring

Work in our laboratory has focused primarily on one of the simplest forms of joint attention, gaze monitoring (GM). Looking at the eyes and abstracting meaning from their movement and expression constitutes one of the important aspects of social-cognitive functioning. Deficits in JA in GM are well documented in autism. These difficulties are present very early in development and constitute one of the key symptoms that differentiate autism spectrum disorder (ASD) from developmental delay in children aged 36 months and younger (Osterling and Dawson 1994; Lord 1995; Baron-Cohen *et al.* 1996; Charman *et al.* 1997; Stone 1997; Volkmar *et al.* 1997; Swettenham *et al.* 1998; Sigman *et al.* 1999). This deficit is relatively stable over time and even older and higher functioning individuals with ASD continue to have marked difficulties monitoring gaze in *unstructured* and *naturalistic* settings and fail to use gaze direction to infer another's desire, goal, or an object of regard (e.g. Baron-Cohen 1989; Baron-Cohen, Baldwin and Crowson 1997; Leekam *et al.* 1997, 1998). Deficits in GM in older children with a verbal mental age of over 4 years

do not appear to have a basis in visual discrimination *per se*, because these individuals do well on tasks designed to test perceptual gaze discrimination explicitly (Baron-Cohen 1989; Leekam *et al*. 1997). The presence of gaze discrimination observed in older and more able individuals (but not younger and lower functioning) might reflect the delayed development of gaze discrimination skills identical to those observed in normal development. But success on a discrimination task could be also achieved through the development or use of alternative mechanisms, as has been reported in behavioural and neuroimaging studies of face perception in high functioning individuals with ASD (Hobson and Lee 1989; Tantam *et al*. 1989; Schultz, Romanski and Tsatsanis 2000a; Critchley *et al*. 2000; Pierce *et al*. 2001). In typical individuals face discrimination relies on configuration or holistic processing and activates the lateral fusiform gyrus area specialised, among other things, in face perception, while the inferior temporal gyrus is activated during tasks requiring object perception.

Typical Development of Gaze Monitoring

Faces and eyes have significance to infants from very early on. Newborn babies preferentially track moving face-like patterns (Johnson and Morton 1991), orientate more frequently to face-like stimuli as compared to non-face-like patterns (Valenza *et al*. 1996) and appear to have some rudimentary ability to discriminate between faces with direct and averted gaze (Farroni 2002). This sensitivity for face-like patterns appears to be mediated by a subcortical neural mechanism that relies primarily on the retinotectal system and which contains a crude specification of the arrangement of the main facial features (Johnson and Morton 1991; Simion *et al*. 1998). It has been hypothesised that this subcortically-mediated preference facilitates detection of human faces in the environment early in ontogenesis (Johnson and Morton 1991). While infants in the first month of life pay more attention to the high-contrast edge area of the face, by 2 months they preferentially scan the region of the eyes (Maurer and Salapatek 1976; Haith, Bergman and Moore 1977; Hainline 1978). The attraction to the eye region in 2 month-old infants is relatively persistent and robust, even when infants are presented with a dynamic image of a 'speaking face', which makes the area of the lips highly competitive from the perceptual point of view (Haith, Bergman and Moore 1977). This finding suggests that the attraction to eyes at the age of 2 months may reflect, for instance, a shift in the status of the face from a mere collection of elements to a meaningful entity, or the emergence of salience of the eyes for social communication. Sensitivity to the eye region of the face does not imply the sensitivity to changes in the gaze direction. It is not until 4–5 months of age that infants are capable of discriminating direction of gaze. Four month-old infants become capable of distinguishing between frontal view photographs with direct and averted eyes

(Vecera and Johnson 1995), and by 5 months infants smile (Lansky and Klein 1979; Hains and Muir 1996; Caron *et al.* 1997) and vocalise more frequently (Bloom 1974) when presented with a face that affords eye contact, as compared with a face that looks slightly to the side or above the infant's line of vision. Although not capable of following the gaze of others spontaneously, 4 month-old infants perceive the movement inherent in gaze shift as a directional cue (Hood, Willen and Driver 1998; Farroni *et al.* 2000). That is, when tested in a spatial cueing attention paradigm, infants have shorter saccadic reaction times to peripheral targets appearing in locations congruent with the eye-gaze direction of the cue (a stimulus face immediately preceding a target presentation) than to targets appearing in incongruent locations (Hood, Willen and Driver 1998; Farroni *et al.* 2000). Eight month-old infants do not usually follow the gaze of another person spontaneously. However, when they are provided with a contingent feedback, they are able to learn that a person's direction of head and gaze shift predicts a location where an interesting event will occur (e.g. a toy will appear from a box on the left if a person in front of the child turns her head and looks in this location; Corkum and Moore 1998). It is not clear whether infants at this age respond to a strong perceptual motion cue inherent in head and gaze shift, as the 4 and 5 month-olds do (Farroni *et al.* 2000), or whether they show some emergent conceptual understanding of the meaning related to this motion. By 10–11 months, infants follow head and gaze turn spontaneously (Corkum and Moore 1998), but it is not until 18–19 months of age that the infants exhibit appreciation of the significance of an eye gaze shift alone (Butterworth and Jarret 1991; Moore and Corkum 1998). It has been hypothesised that onset of gaze monitoring, which is relatively frequent and independent from the presence or absence of the targets in the visual field, at this age represents a conceptual shift toward understanding that gaze shift signifies attentional shift of focus of another person (Moore and Corkum 1998). The conceptual understanding of attentional significance of gaze continues to develop well into preschool age (e.g. Anderson and Doherty 1997).

Neural Bases of Gaze Processing

Recent neuroimaging, neurophysiological and behavioural studies suggest that mechanisms underlying the perception of eye-gaze direction are complex. The ability to discriminate changes in eye gaze direction depends on the functioning of a high-level visual processing area in the temporal cortex that receives polysensory input, the superior temporal sulcus (Allison, Puce and McCarthy 2000). In humans, this area is primarily concerned with the perception of static and moving facial components, (e.g. eyes and mouth), head orientation and movement, as well as goal-orientated body movements (Perrett *et al.* 1985; Hasselmo, Rolls and Baylis 1989; Campbell *et al.* 1990; Grafton *et al.* 1996;

Rizzolatti *et al*. 1996; Puce *et al*. 1998; Wicker *et al*. 1998; Hoffman and Haxby 2000). The superior temporal sulcus as a neural structure involved in gaze perception does not function, however, like an 'encapsulated module' (Baron-Cohen 1994), but rather works in interaction with the lateralised temporal lobe subsystems specialised for processing upright faces, the fusiform gyrus (e.g. Vecera and Johnson 1995; Puce *et al*. 1998; Kingstone, Friesen and Gazzaniga 2000). Moreover, perception of the eyes in some contexts activates the amygdala, a structure that plays a role in reading social signals from the face (Kawashima *et al*. 1999).

Implications for Studies on Gaze Abnormalities in ASD

The ontogenesis of eye gaze monitoring and understanding that visual attention links others to the external world is a prolonged and complex process in which innate predispositions interact with a range of perceptual, attentional, social and cognitive factors. Although deficit in spontaneous gaze monitoring is widely recognised as an early sign of autism, the processes underlying this deficit remain largely unknown. It is not clear whether the failure to monitor gaze observed in individuals with an ASD results from perceptual deficits or a failure to translate the perceptual knowledge about gaze into conceptual understanding of the significance of gaze as an index of the intentions and goals of others. Documenting specific to ASD deficits in the forerunners of gaze monitoring, as well as defining the point at which the developmental trajectories of autistic, developmentally delayed, and typically developing children diverge, would be of great theoretical and practical significance. First, it may contribute to the designing of new behaviour-based diagnostic instruments for autism in children prior to 18 months of age. Considering a 5–10% recurrence rate for the broader phenotype siblings of children with autism (Rutter *et al*. 1997), such methods might be useful in identifying autism in the first year of life. Second, it may provide suggestions for behaviours that need to be targeted for intervention at this early age, which is crucial considering the high neuroplasticity present in early ontogenesis (e.g. Johnson 1999). Third, it may generate new hypotheses regarding the mechanisms and neural structures responsible for social eye gaze behaviour deficits in autism.

FACE PERCEPTION IN AUTISM

One of the most striking areas of disability in autism are the problems that individuals with autism have in face-to-face social engagement; aversion of eye contact, difficulties in mutual gaze and pragmatic communication, and difficulties in face perception have been described repeatedly (Grossman, Carter and Volkmar 1997). Whereas most people qualify as face experts

(Tanaka and Gauthier 1997), individuals with autism and related conditions are selectively impaired in their ability to recognise faces (Langdell 1978; Hobson 1986a,b; Weeks and Hobson 1987; Braverman *et al.* 1989; Hobson and Lee 1989; MacDonald *et al.* 1989; Tantam *et al.* 1989; Szatmari *et al.* 1990; Boucher and Lewis 1992; Davies and Bishop 1994; Klin *et al.* 1999). Behavioural data suggest that persons with autism approach the problem of facial identity recognition in a more piecemeal fashion, and rely heavily on analysis of individual face features rather than configuration of features (Langdell 1978; Hobson, Ouston and Lee 1988; Tantam *et al.* 1989; Joseph 2001; Klin *et al.* 2002a,b). One interpretation is that persons with an ASD fail to develop expertise for faces because of inadequate attention to faces across development (Schultz, Romanski and Tsatsanis 2000; Schultz *et al.* 2000; Grelotti, Gauthier and Schultz 2002).

Research during the last two decades using several different methods have made it abundantly clear that in typically developing individuals there is a region on the underside of the temporal lobe in the middle portion of the fusiform gyrus, especially on the right, that is more important for face perception than other types of (non-expert) object perception (Damasio, Damasio and Van Hoesen 1982; Allison *et al.* 1994; Haxby *et al.* 1994, 1999; Kanwisher, McDermott and Chun 1997). In a recent literature review, Kanwisher (2000) notes that fusiform face area activity is at least twice as strong to faces as to a wide range of non-face stimuli, such as assorted objects, animals without heads, and the backs of human heads.

Our group has reported (Schultz, Romanski and Tsatsanis 2000) that persons with autism or Asperger's syndrome have significantly less fMRI activity in the middle portion of the right fusiform gyrus when performing face identity discrimination tasks. This finding has now been replicated (e.g. Critchley *et al.* 2000).

Whereas the fusiform gyrus is important for the perception of facial identity (Kanwisher 2000), a medial temporal lobe structure, the amygdala, has been shown to play a critical role in the early stage processing of facial expression (Breiter *et al.* 1996; Morris *et al.* 1996). The amygdala appears to be a fast-responding structure that quickly interprets emotionally potent stimuli (LeDoux 1995). When fearful faces are presented rapidly (< 30 ms), below the subject's threshold of conscious awareness (i.e. without fusiform involvement), the amgydala nevertheless detects these facial displays of affect (Whalen *et al.* 1998). Thus, early on, identity and expression appear to be processed separately, although downstream information processes may converge (Schweinberger and Soukup 1998).

In addition to the early role of the amygdala, good evidence exists to suggest that posterior aspects of the superior temporal sulcus and the amygdala are involved in reading all types of dynamic non-verbal communications of affect, including emotional expression, gesture, and social cues provided by

interpreting the direction of eye gaze (Puce *et al*. 1998; Allison *et al*. 1999; Critchley *et al*. 2000). Two published reports suggest that the amygdala and the superior temporal sulcus are underactive in ASDs, in both explicit (Ring *et al*. 1999) and implicit perceptual tasks involving facial expression (Critchley *et al*. 2000).

The general ability of typically developing infants to almost instantaneously recognise familiar faces represents an impressive feat in visual recognition which has led some researchers to suggest that essentially all normal adults are *experts* in the recognition of faces (Tanaka and Gauthier 1997). As noted previously, behavioural studies suggest that the process of face recognition in autism may be strikingly different and the effect seems largely to be confined to faces, since non-face object recognition seems to be intact (Hobson, Ouston and Lee 1988; Tantam *et al*. 1989; Boucher and Lewis 1992; Davies and Bishop 1994). For example, Langdell (1978) noted that children with autism were better than controls at processing inverted faces, and our own work (Schultz, Romanski and Tsatsanis 2000) tends to support the notion that faces are processed in terms of features rather than overall configuration. These findings underscore the importance of developing novel approaches in which specific *processes* can be identified and then related to brain function. One might also hope that such approaches would clarify precise mechanisms and pathways to dysfunction, e.g. in contrast to the present more global and 'scatter-shot' approaches employed in assessment of social competence. Such approaches might also have important implications in other areas, e.g. more quantifiable indices of specific aspects of social competence might also be used in the study of the broader autism spectrum phenotype in family members (Rutter *et al*. 1997).

Eye Tracking in Autism

One of the limitations of the body of work on facial perception in autism has been the reliance on highly artificial and unnaturalistic experimental methods. For example, tasks have involved examination of still faces presented out of any social context. In an effort to provide a more ecologically valid and process-orientated approach, we (Klin *et al*. 2002a) have recently developed an experimental paradigm to measure social functioning using more naturalistic stimuli. This approach, in which small and unobtrusive video cameras capture the ongoing point of regard of the observer, allows us to see the world through the eyes of an individual with autism. With this approach, the point of regard is superimposed on the ongoing images of a videotaped scene which is then available for analysis of viewing patterns.

The potential utility of this approach is captured in a series of illustrations in which the viewing patterns of a high functioning individual with autism (full-scale IQ 119) is contrasted to an age-, gender- and IQ-matched control in

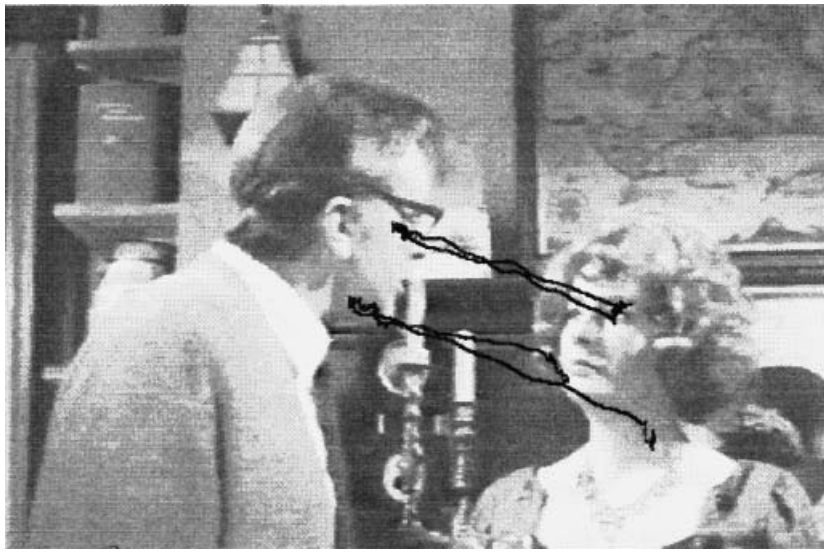


Figure 9.2. Visual focus of an autistic man (lower trace) and a normal comparison subject (upper trace) showing a film clip of a conversation. Reprinted from Klin *et al.* (2002b)

watching a series of digitised clips of the movie version of *Who's Afraid of Virginia Woolf?*—a movie chosen because of its content-rich social situations, likely to maximise viewers' monitoring of each person's socially expressive actions as well as those characters' reactions to the actions of others. The demanding social complexity in the movie was intended to mirror complicated social situations. Examination of the data from this initial report vividly illustrates the marked differences that these two individuals utilised in attempting to extract meaning from the videoclips (Klin *et al*. 2002a).

Consistent with previous literature (Langdell 1978; Hobson *et al*. 1988a; Happé 1994) the individual with autism relied much more heavily on mouths rather than eyes. Figure 9.2 presents 2 s of data showing eye movement sumperimposed onto a single still frame; in the ongoing scene there is affectively laden discussion between two of characters. The typical viewer (upper trace) shifts from eye to eye, while the individual with autism (lower trace) focuses on the mouths and adjacent regions. The over-reliance on the mouth is also shown in Figure 9.3, in which the character portrayed by Elizabeth Taylor is silent for over 13 s; although this segment is silent, there is growing discomfort in the scene which is conveyed largely by the actress's gaze. In Figure 9.3 the eye-tracking data are collapsed onto one still frame. Again, the viewer with autism (lower trace) is entirely focused on the mouth and lower face, while the typical viewer (upper trace) focuses primarily on the eyes.

Figure 9.3. Visual focus of an autistic man (lower trace) and a normal comparison subject (upper trace) shown a film clip of a silent actor. Reprinted from Klin *et al.* (2002b)

The data from the original case study (Klin *et al.* 2002a) also highlighted the difficulty the high functioning man had in interpretation of non-verbal affective cues, i.e. his tendency to focus only on the speaker, and even then only on the speaker's mouth, meant that he was not able to process any of the important affective and other non-verbal information provided by the non-speaking actor, neither, for that matter, was he able to attend to any of the cues provided by eyes and upper face of the person who was speaking. This was vividly illustrated during clips in which one character was embarrassing her husband — the husband's many non-verbal cues as to his growing discomfort and embarrassment were entirely lost to the higher-functioning individual with autism, who attended only to the speech of the wife — the only character talking.

The difficulties in negotiating the microcosm of the social world presented in these clips was also illustrated by the response of the participants to a highly charged scene in which the character portrayed by Elizabeth Taylor is seducing the younger man in the presence of her husband, Richard Burton, who stands in the background. As can be seen in Figure 9.4, the typical viewer actively scans the loaded social triangle presented in this 7 s clip, while the viewer with autism does not perceive the importance of the husband or the centrality of the three-way interaction. On the other hand, in analysis of other segments of

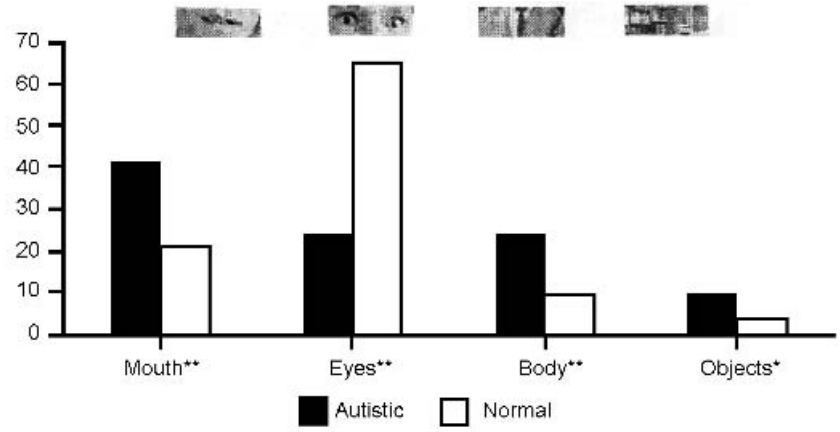


Figure 9.4. Percentage of viewing time spent focused on mouth, eyes, body and object regions. Group data adapted from Klin *et al.* 2002b

the movie, in which physical rather than social cues are predominant, the individual with autism did not display deficits.

Data from a larger series of cases contrasting a group of higher functioning individuals with autism to typically developing age- and IQ-matched controls (Klin *et al*. 2002b) confirms the robustness of these findings. A group of 15 adolescent and adult males with high functioning autism were age- and IQ-matched to typically developing controls and shown a series of videoclips. For the purposes of analysis, scanning patterns were predefined to include mouth, eye, body and object regions. As shown in Figure 9.4, there were dramatic and significant differences for each viewing region.

The figures underscore the importance of focusing on specific processes that may underlie social dysfunction in autism. The pronounced focus on mouths and tendency to attend to physical rather than social cues are vividly illustrated in these data, since the individuals with autism were more than twice as likely to focus on mouths and more than twice as likely *not* to focus on the eye region of face. These observations are also striking given the early-emerging tendency, in typically developing infants, to focus on eyes and social stimuli (Baron-Cohen, Wheelwright and Jolliffe 1997; Dawson *et al*. 1998). The extension of this work to younger and lower functioning individuals will be important.

SUMMARY

Autism is a neurodevelopmental disorder of early onset marked by a profound social disability affecting a person's capacity for understanding other people,

intuiting their feelings, and establishing reciprocal relationships. Although the social disability of autism has remained a central defining feature of the condition, the mechanisms underlying this disability remain poorly understood. In this chapter we have reviewed the nature of the social deficit and some of the theories proposed to account for it as well as current research on specific mechanisms. One unifying theme has been the attempt to more precisely specify this highly heterogeneous social phenotype. Improved methods of characterisation of the social deficit, with a closer approximation of experimental methods to the naturalistic demands of real-life social situations, may facilitate the development of a more comprehensive neurobiological theory. Recent work has focused on differences in processing of social stimuli (recognition of faces vs. objects) in the cerebral cortex, and differences in the relative salience of different aspects of social cues in more naturalistic settings, and in differences in joint attention which appear to be of very early onset. This research effort builds on an emerging synergy of different branches of social neuroscience.

Future work must emphasise multiple, interdisciplinary perspectives in understanding the social difficulties of individuals with autism. It will need to be conducted at various levels of analysis from the level of the genetic substrate of social adaptation to studies of the perception of social-affective signals and face stimuli, to studies of the treatment of the disabling social anxiety that frequently is associated with these conditions.

REFERENCES

Adolphs, R., Damasio, H., Tranel, D. and Damasio, A.R. (1996). Cortical systems for the recognition of emotion in facial expressions. *J. Neurosci.* **16**, 7678–7687.

Allison, T., McCarthy, G., Nobre, A.C., Puce, A. and Belger A. (1994). Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. *Cerebr. Cortex* **5**, 544–554.

Allison, T., Puce, A., Spencer, D.D. and McCarthy, G. (1999) Electrophysiological studies of human face perception. I. Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebr. Cortex* **9** (5), 415–430.

Allison, T., Puce, A. and McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.* **4**, 267–278.

Alsobrook II, J.P., Leckman, J.F., Goodman, W.K., Rasmussen, S.A. and Pauls, D.L. (1999). Segregation analysis of obsessive-compulsive disorder using symptom-based factor scores. *Am. J. Med. Genet.* **88**, 669–675.

APA (American Psychiatric Association) (1980). *Diagnostic and Statistical Manual*. APA Press, Washington, DC.

APA (American Psychiatric Association) (1994). *Diagnostic and Statistical Manual*, 4th edn. APA Press, Washington, DC.

Anderson, J.R. and Doherty, M. (1997). Preschoolers' perception of other people's looking: photographs and drawings. *Perception* **26**, 333–343.

Anderson, G.M. and Hoshiono, Y. (1997). Neurochemical studies of autism. In D.J. Cohen and F.R. Volkmar (eds), *Handbook of Autism and Pervasive Developmental Disorders*. Wiley, New York, pp. 325–343.

Bachevalier, J. (1996). Medial temporal lobe and autism: a putative animal model in primates. *J. Autism Dev. Disord.* **26**, 217–220.

Bachevalier, J., Malkova, L. and Mishkin, M. (2001). Effects of selective neonatal temporal lobe lesions on socioemotional behavior in infant rhesus monkeys (*Macaca mulatta*). *Behav. Neurosci.* **115**, 545–559.

Baldwin, D.A. (1995). Understanding the link between joint attention and language. In C. Moore and P. Dunham (eds), *Joint Attention: Its Origins and Role in Development*. Erlbaum, Hillsdale, NJ, pp. 103–130.

Baron-Cohen, S. (1988). Social and pragmatic deficits in autism: cognitive or affective? *J. Autism Dev. Disord.* **18**, 379–402.

Baron-Cohen, S. (1989). Perceptual role taking and protodeclarative pointing in autism. *Br. J. Dev. Psychol.* **7**, 113–127.

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

Baron-Cohen, S. (1994). How to build a baby that can read minds: cognitive mechanisms in mind reading. *Cahiers Psychol. Cogn. [Current Psychology of Cognition]* **13**, 513–532.

Baron-Cohen, S., Baldwin, D. and Crowson, M. (1997). Do children with autism use the speaker's direction of gaze strategy to crack the code of language? *Child Dev.* **68**, 48–57.

Baron-Cohen, S., Ring, H., Moriarty, J., Schmitz, B. *et al.* (1994). Recognition of mental state terms: clinical findings in children with autism and a functional neuroimaging study of normal adults. *Br. J. Psychiat.* **165**, 640–649.

Baron-Cohen, S., Campbell, R., Karmiloff-Smith, A., Grant, J. and Walker, J. (1996). Are children with autism blind to the mental significance of the eyes? *Br. J. Dev. Psychol.* **4**, 124–145.

Baron-Cohen, S., Wheelwright, S. and Jolliffe, T. (1997). Is there a 'language of the eyes'? Evidence from normal adults, and adults with autism or Asperger's syndrome. *Visual Cogn.* **4**, 311–331.

Bates, E. (1979). On the evolution and development of symbols. In E. Bates (ed.), *The Emergence of Symbols: Cognition and Communication in Infancy*. Academic Press, New York, pp. 1–32.

Bloom, K. (1974). Eye contact as a setting event for infant learning. *J. Exp. Psychol.* **17**, 250–263.

Boucher, J. and Lewis, V. (1992). Unfamiliar face recognition in relatively able autistic children. *J. Child Psychol. Psychiat.* **33**, 843–859.

Bowler, D.M. (1992). 'Theory of Mind' in Asperger's syndrome. *J. Child Psychol. Psychiat.* **33**, 877–893.

Braverman M., Fein, D., Lucci, D. and Waterhouse, L. (1989). Affect comprehension in children with pervasive developmental disorders. *J. Autism Dev. Disord.* **19**, 301–316.

Breiter, H.C., Etcoff, N.L., Whalen, P.J., Kennedy, W.A. *et al.* (1996). Response and habituation of human amygdala during visual processing of facial expression. *Neuron* **17**, 875–887.

Brothers, L., Ring, B. and Kling, A. (1990). Response of neurons in the macaque amygdala to complex social stimuli. *Behav. Brain Res.* **41**, 199–213.

Buitelaar, J.K., van Engeland, H., de Kogel, C.H., de Vries, H. *et al.* (1992). Deficits in social behavior in autism and their modification by a synthetic adrenocorticotrophic hormone (4-9) analog. *Experientia* **48**, 391–394.

Buitelaar, J.K., van Engeland, H., de Kogel, K.H., de Vries, H. and van Heof, J.A. (1991). Differences in the structure of social behaviour of autistic children and non-autistic retarded controls. *J. Child Psychol. Psychiat.* **32**, 995–1015.

Butterworth, G. and Jarrett, N. (1991). What minds have in common is space: spatial mechanisms serving joint visual attention in infancy. *Br. J. Dev. Psychol.* **9**, 55–72.

Butterworth, G. (1995). Origins of mind in perception and action. In C. Moore and P. Dunham (eds), *Joint Attention: Its Origins and Role in Development*. Erlbaum, Hillsdale, NJ, pp. 103–130.

Campbell, R., Heywood, C.A., Cowey, A., Regard, M. and Landis, T. (1990). Sensitivity to eye gaze in prosopagnosic patients and monkeys with superior temporal sulcus ablation. *Neuropsychologia* **28**, 1123–1142.

Capps, L., Sigman, M., Mundy, P. and Neumann, C.G. (1994). Attachment security in children with autism. *Dev. Psychopathol.* **6**, 249–261.

Caron, A.J., Caron, R., Roberts, J. and Brooks, R. (1997). Infant sensitivity to deviation in dynamic facial–vocal displays: the role of eye regard. *Dev. Psychol.* **33**, 802–813.

Castelli, F., Frith C., Happé, F. and Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain* **125**, 1839–1849.

Charman, T., Baron-Cohen, S., Swettenham, J., Cox, A. *et al.* (1997). Infants with autism: an investigation of empathy, pretend play, joint attention and imitation. *Dev. Psychol.* **33**, 781–789.

Chawarska, K., Volkmar, F.R. and Klin, A. (2002). Gaze perception in young children with autism. Paper presented at the American Academy of Child and Adolescent Psychiatry, San Francisco, CA, 26 October.

Constantino, J.N., Przybeck, T., Friesen, D. and Todd, R.D. (2000). Reciprocal social behavior in children with and without pervasive developmental disorders. *J. Dev. Behav. Pediat.*, **21**, 2–11.

Corkum, V. and Moore, C. (1998). The origins of joint attention in infants. *Dev. Psychol.* **34**, 28–38.

Courchesne, E. (1997). Brainstem, cerebellar and limbic neuroanatomical abnormalities in autism [published erratum appears in *Curr. Opin. Neurobiol.* 1997, **7**, 568]. *Curr. Opin. Neurobiol.* **7**, 269–278.

Courchesne, E., Yeung-Courchesne, R., Press, G.A., Hesselink, J.R. and Jernigan, T.L. (1988). Hypoplasia of cerebellar vermal lobules VI and VII in autism. *N. Engl. J. Med.* **318**, 1349–1354.

Critchley, H.D., Daly, E.M., Bullmore, E.T., Williams, S.C. *et al.* (2000). The functional neuroanatomy of social behaviour: changes in cerebral blood flow when people with autistic disorder process facial expressions. *Brain* **123**, 2203–2212.

Dahlgren, S.O. and Trillingsgaard, A. (1996). Theory of mind in non-retarded children with autism and Asperger's syndrome. A research note. *J. Child Psychol. Psychiat.* **37**, 759–763.

Damasio, A.R. (1995). On some functions of the human prefrontal cortex. *Ann. N. Y. Acad. Sci.* **769**, 241–251.

Damasio, A. (2001). Fundamental feelings. *Nature* **413** (6858), 781.

Damasio, A.F., Damasio, H. and Van Hoesen, G.W. (1982). Prosopagnosia: anatomical basis and behavioral mechanisms. *Neurology* **32**, 331–341.

Davies, S. and Bishop, D. (1994). Face perception in children with autism and Asperger's syndrome. *J. Child Psychol. Psychiat.* **35**, 1033–1057.

Dawson, G., Meltzoff, A.N., Osterling, J. and Rinaldi, J. (1998). Neuropsychological correlates of early symptoms of autism. *Child Dev*. **69**, 1276–1285.

Farroni, T., Johnson, M.H., Brockbank, M. and Simion, F (2000). Infants' use of gaze direction to cue attention: the importance of perceived motion. *Vis. Cognit*. **7**, 705–718.

Farroni, R., Csibra, G., Simion, F. and Johnson, M. (2002). Eye contact detection in humans from birth. *Proc. Natl Acad. Sci. USA* **99**, 9602–9695.

Fletcher, P.C., Happe, F., Frith, U., Baker, S.C. *et al*. (1995). Other minds in the brain: a functional imaging study of 'theory of mind' in story comprehension. *Cognition* **57**, 109–128.

Fombonne, E. (1999). The epidemiology of autism: a review. *Psychol. Med*. **29**, 769–786.

Freeman, B.J., Rahbar, B., Ritvo, E., Bice, T.L. *et al*. (1991). The stability of cognitive and behavioral parameters in autism: a 12 year prospective study. *J. Am. Acad. Child Adolesc. Psychiat*. **30**, 479–482.

Frith, U. (1989). *Autism: Explaining the Enigma*. Oxford: Basil Blackwell.

Gauthier, I., Tarr, M., Anderson, A., Skudlarski, P. and Gore, J. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neurosci*. **2**, 568–573.

Gillberg, C., Ehlers, S., Schaumann, H., Jakobsson, G. *et al*. (1990). Autism under age 3 years: a clinical study of 28 cases referred for autistic symptoms in infancy [see comments]. *J. Child Psychol. Psychiat*. **31**, 921–934.

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

Grafton, S.T., Arbib, M.A., Fadiga, L. and Rizzolatti, G. (1996). Localization of grasp representations in humans by PET 1: observation compared with imagination. *Exp. Brain Res*. **112**, 103–111.

Grigorenko, E.L., Wood, F.B., Meyer, M.S., Hart, L.A. *et al*. (1997). Susceptibility loci for distinct components of developmental dyslexia chromosomes 6 and 15 [see comments]. *Am. J. Hum. Genet*. **60**, 27–39.

Grossman, J.B., Carter, A. and Volkmar, F.R. (1997). Social behavior in autism. *Ann. N. Y. Acad. Sci*. **807**, 440–454.

Guillaume, P. (1971). *Imitation in Children*. University of Chicago Press, Chicago.

Hainline, L. (1978). Developmental changes in visual scanning of face and non-face patterns by infants. *J. Exp. Child Psychol*. **25**, 90–115.

Hains, S.M.J. and Muir, D.W. (1996). Infant sensitivity to adult eye direction. *Child Dev*. **67**, 190–1951.

Haith, M.M., Bergman, T. and Moore, M. (1977). Eye contact and face scanning in early infancy. *Science* **218**, 179–181.

Happé, F., Ehlers, S., Fletcher, P., Frith, U. *et al*. (1996). 'Theory of mind' in the brain: evidence from a PET scan study of Asperger's syndrome. *NeuroReport* **8**, 197–201.

Happé, F.G. (1994). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *J. Autism Dev. Disord*. **24**, 129–154.

Hasselmo, M.E., Rolls, E.T. and Baylis, G.C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behav. Brain Res*. **32**, 203–218.

Haxby, J.V., Horwitz, B., Ungerleider, L.G., Maisog, J. *et al*. (1994). The functional organization of human extrastriate cortext: a PET-rCBFstudy of selective attention to faces and locations. *J. Neurosci*. **14**, 6336–6353.

Haxby, J.V., Ungerleider, L.G., Clark, V.P., Schouten, J.L. *et al*. (1999). The effect of face inversion on activity in human neural systems for face and object perception. *Neuron* **22**, 189–199.

Hobson, R.P. and Lee, A. (1989). Emotion-related and abstract concepts in autistic people: evidence from the British Picture Vocabulary Scale. *J. Autism Dev. Disord*. **19**, 601–623.

Hobson, R.P. and Lee, A. (1998). Hello and goodbye: a study of social engagement in autism. *J. Autism Dev. Disord*. **28**, 117–127.

Hobson, R.P. (1986a). The autistic child's appraisal of expressions of emotion. *J. Child Psychol. Psychiat*. **27**, 321–342.

Hobson, R.P. (1986b). The autistic child's appraisal of expressions of emotion: a further study. *J. Child Psychol. Psychiat*. **27**, 671–80.

Hobson, R.P., Ouston, J. and Lee, A. (1988). Emotion recognition in autism: coordinating faces and voices. *Psycholog. Med*. **18**, 911–923.

Hoffman, E. and Haxby, J.V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neurosci*. **3**, 80–84.

Hood, B.M., Willen, J.D. and Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychol. Sci*. **9**, 131–134.

Hughes, C., Leboyer, M. and Bouvard, M. (1997). Executive function in parents of children with autism. *Psychol. Med*. **27**, 209–220.

Hughes, C., Plumet, M.H. and Leboyer, M. (1999). Towards a cognitive phenotype for autism: increased prevalence of executive dysfunction and superior spatial span amongst siblings of children with autism. *J. Child Psychol. Psychiat. Allied Discip*. **40**, 705–718.

IMGSAC (International Molecular Genetic Study of Autism Consortium) (1998). A full-genome screen for autism with evidence for linkage to a region on chromosome 7q. *Hum. Mol. Genet*. **7**, 571–578.

Johnson, M.H. (1999). Cortical plasticity in normal and abnormal cognitive development: evidence and working hypotheses. *Dev. Psychopathol*. **11**, 419–437.

Johnson, M.H. and Morton, J. (1991). *Biology and Cognitive Development: the Case of Face Recognition*. Basil Blackwell, Oxford.

Joseph, R.M. (2001). *Face Recognition Processes in Typically Developing Children and Children with Autism*. Society for Research in Child Development, Minneapolis, MN.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nerv. Child* **2**, 217–250.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neurosci*. **3**, 759–763.

Kanwisher, N., McDermott, J. and Chun, M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci*. **17**, 4302–4311.

Kawashima, R., Sugiura, M., Kato, T., Nakamura, A. *et al*. (1999). The human amygdala plays an important role in gaze monitoring. A PET study. *Brain* **122**, 779–783.

Kemper, T. L. and Bauman, M. (1998). Neuropathology of infantile autism. *J. Neuropathol. Exp. Neurol*. **57**, 645–652.

Kingstone, A., Friesen, C. and Gazzaniga, M.S. (2000). Reflexive joint attention depends on lateralized cortical connections. *Psychol. Sci*. **11**, 159–166

Klin, A., Jones, W., Schultz, R., Volkmar, F.R. and Cohen, D.J. (2002a). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Am. J. Psychiat*. **59**, 809–816.

Klin, A., Jones, W., Schultz, R., Volkmar, F. and Cohen, D. (2002b). Defining and qualifying the social phenotype in autism. *Am. J. Psychiat*. **159**, 895–908.

Klin, A., Schultz, R. and Cohen, D. (2000). Theory of mind in action: developmental perspectives on social neuroscience. In S. Baron-Cohen, H. Tager-Flusberg and D. Cohen (eds), *Understanding Other Minds: Perspectives from Developmental Neuroscience*, 2nd edn. Oxford University Press, Oxford, pp. 357–388.

Klin, A., Volkmar, F.R. and Sparrow, S.S. (1992). Autistic social dysfunctions: some limitations of the theory of mind hypothesis. *J. Child Psychol. Psychiat.* **33**, 861–876.

Klin, A., Sparrow, S.S., de Bildt, A., Cicchetti, D.V. *et al.* (1999). A normed study of face recognition in autism and related disorders. *J. Autism Dev. Disord.* **29**, 497–507.

Lane, R.D., Reiman, E.M., Ahern, G.L., Schwartz, G.E. and Davidson, R.J. (1997a). Neuroanatomical correlates of happiness, sadness, and disgust. *Am. J. Psychiat.* **154**, 926–933.

Lane, R.D., Reiman, E.M., Bradley, M.M., Lang, P.J. *et al.* (1997b). Neuroanatomical correlates of pleasant and unpleasant emotion. *Neuropsychologia* **35**, 1437–1444.

Langdell, T. (1978). Recognition of faces: an approach to the study of autism. *J. Child Psychol. Psychiat.* **19**, 255–268.

Lansky, R.E. and Klein, R.E. (1979). The reactions of 5 month-olds to eye contact of the mother and of a stranger. *Merrill–Palmer Qu.* **25**, 163–170.

LeDoux, J.E. (1995). *The Emotional Brain*. Simon and Schuster, New York.

Leekam, S.R., Lopez, B. and Moore, C. (2000). Attention and joint attention in preschool children with autism. *Dev. Psychol.* **36**, 261–273.

Leekam, S.R., Baron-Cohen, S., Perrett, D., Milders, M. and Brown, S. (1997). Eye-direction detection: a dissociation between geometric and joint attention skills in autism. *Br. J. Dev. Psychol.* **15**, 77–95.

Leekam, S.R., Hunnisett, E. and Moore, C. (1998). Target and cues: gaze-following in children with autism. *J. Child Psychol. Psychiat.* **39**, 951–962.

Lord, C. (1997). Diagnostic instruments in autism spectrum disorders. In D.J. Cohen and F.R. Volkmar (eds), *Handbook of Autism and Pervasive Developmental Disorders*, 2nd edn. Wiley, New York, pp. 460–483.

Lord, C. and Paul, R. (1997). Language and communication in autism. In D.J. Cohen and F.R. Volkmar (eds), *Handbook of Autism and Pervasive Developmental Disorders*, 2nd edn. Wiley, New York, pp. 195–225.

Lord, C. (1995). Follow up of 2 year-old referred for possible autism. *J. Child Psychol. Psychiat.* **36**, 1365–1382

Loveland, K.A. and Kelley, M.L. (1991). Development of adaptive behavior in preschoolers with autism or Down syndrome. *Am. J. Ment. Retard.* **96**, 13–20.

MacDonald, H., Rutter, M., Howlin, P., Le Conteur, A. *et al.* (1989). Recognition and expression of emotional cues by autistic and normal adults. *J. Child Psychol. Psychiat.* **30**, 865–877.

Marcus, D.J. and Nelson, C.A. (2001). Neural bases and development of face recognition in autism. *CNS Spectr.* **6**, 36–58.

Maurer, D. and Salapatek, P. (1976). Developmental changes in the scanning of faces by young infants. *Child Dev.* **47**, 523–527.

Minshew, N.J., Sweeney, J.A. and Bauman, M.L. (1997). Neurological aspects of autism. In D.J. Cohen and F.R. Volkmar (eds), *Handbook of Autism and Pervasive Developmental Disorders*, 2nd edn. Wiley, New York, pp. 344–369.

Morisette, P., Ricard, M. and Decarie, T. (1995). Interactions between shared attention during the preverbal period and the development of verbal references. *Int. J. Psychol.* **30**, 481–498.

Morris, J.S., Frith, C.D., Perrett, D.I., Rowland, D. *et al.* (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature* **383**, 812–815.

Mottron, L., Peretz, I. and Menard, E. (2000). Local and global processing of music in high-functioning persons with autism: beyond central coherence? *J. Child Psychol. Psychiat*. **41**, 1057–1065.

Moore, C. and Corkum, V. (1998). Infant gaze following based on eye direction. *Br. J. Dev. Psychol*. **16**, 495–503.

Moore, C. (1999). Gaze following and the control of attention. In P. Rochat (ed.), *Early Social Cognition*. Erlbaum, Mahwah, NJ, pp. 241–256.

Mundy, P. (1995). Joint attention and social-emotional approach behavior in children with autism. Special Issue: Emotions in Developmental Psychopathology. *Dev. Psychopathol*. **7**, 63–82.

Mundy, P. and Crowson, M. (1997). Joint attention and early social communication: implications for research on intervention with autism. *J. Autism Dev. Disord*. **27**, 653–676.

Mundy, P., Sigman, M., Ungerer J. and Sherman, T. (1986). Defining the social deficits of autism: the contribution of non-verbal communication measures. *J. Child Psychol. Psychiat*. **27**, 657–669.

Mundy, P. and Gomes, A. (1998). Individual differences in joint attention skill development in the second year of life. *Infant Behavior and Development* **21**, 469–482.

Murphy, C.M. and Messer, D.J. (1977). Mothers, infants, and pointing: a study of gesture. In H.R. Schaffers (ed.), *Studies in Mother–Infant Interaction*. Academic Press, London, pp. 325–354.

National Academy of Sciences (2001). *Educating Young Children with Autism*. NAS Press, Washington, DC.

O'Neill, M. and Jones, R.S.P. (1997). Sensory-perceptual abnormalities in autism: a case for more research? *J. Autism Dev. Disord*. **27**, 283–293.

Ornitz, E.M., Guthrie, D. and Farley, A.J. (1977). The early development of autistic children. *J. Autism Childhood Schizophren*. **7**, 207–229.

Osterling, J. and Dawson, G. (1994). Early recognition of children with autism: a study of first birthday home videotapes. *J. Autism Dev. Disord*. **24**, 247–257.

Osterling, J. and Dawson, G. (1994). Early recognition of children with autism: a study of first birthday home videotapes. *J. Autism Dev. Disord*. **24**, 247–257.

Ozonoff, S. (1997). Causal mechanisms of autism: unifying perspectives from an information-processing framework. In D.J. Cohen and F.R. Volkmar (eds), *Handbook of Autism and Pervasive Developmental Disorders*, 2nd edn. Wiley, New York, pp. 868–879.

Ozonoff, S., Pennington, B.F. and Rogers, S.J. (1991). Executive function deficits in high-functioning autistic individuals: relationship to theory of mind. *J. Child Psychol. Psychiat*. **32**, 1081–1105.

Pennington, B.F. and Ozonoff, S. (1996). Executive functions and developmental psychopathology. *J. Child Psychol. Psychiat*. **37**, 51–87.

Perrett, D.I., Smith, P.A.J., Potter, D.D., Mistlin, A.J. *et al*. (1985). Visual analysis of body movement by neurons in the temporal cortex of the macaque monkey: a preliminary report. *Behav. Brain Res*. **16**, 153–177.

Pickles, A., Starr, E., Kazak, S., Bolton, P. *et al*. (2000). Variable expression of the autism broader phenotype: findings from extended pedigrees. *J. Child Psychol. Psychiat. Allied Discipl*. **41**, 491–502.

Pierce, K., Muller, R.A., Ambrose, J., Allen, G. and Courchesne, E. (2001). Face processing occurs outside the fusiform 'face area' in autism: evidence from functional MRI. *Brain* **124**, 2059–2073.

Piven, J., Nehme, E., Piven, J., Nehme, E., Simon, J., Barta, P. *et al*. (1992). Magnetic resonance imaging in autism: measurement of the cerebellum, pons, and fourth ventricle. *Biol. Psychiat*. **31**, 491–504.

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

Ring, H.A., Baron-Cohen, S., Wheelwright, S., William, S.C. *et al*. (1999). Cerebral correlates of preserved cognitive skills in autism: a MRI study of embedded figures task performance. *Brain* **122**, 1305–1315.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V. *et al*. (1996). Localization of grasp representations in humans by PET 1: observation vs. execution. *Exp. Brain Res*. **111**, 246–252.

Rodrigue, J.R., Morgan, S.B. and Geffken, G.R. (1991). A comparative evaluation of adaptive behavior in children and adolescents with autism, Down's syndrome, and normal development. *J. Autism Dev. Disord*. **21**, 187–196.

Rutter, M. (1978). Diagnosis and definition of childhood autism. *J. Autism Child Schizophr*. **8**, 139–161.

Rutter, M., Bailey, A., Simonoff, E. and Pickles, A. (1997). Genetic influences in autism. In D.J. Cohen and F.R. Volkmar (eds), *Handbook of Autism and Pervasive Developmental Disorders*, 2nd edn. Wiley, New York, pp. 370–387.

Schultz, R.T., Gauthier, I., Klin, A., Fulbright, R. *et al*. (2000). Abnormal ventral temporal cortical activity among individuals with autism and Asperger's syndrome during face discrimination. *Arch. Gen. Psychiat*. **57**, 331–340.

Schultz, R.T., Grelotti, D.J., Klin, A., Kleinman, J. *et al*. (in press). The role of the fusiform face area in social cognition: implications for the pathobiology of autism. *Phil. Trans. R. Soc. Ser. B*.

Schultz, R.T., Romanski, L.M. and Tsatsanis, K.D. (2000). Neurofunctional models of autistic disorder and Asperger's syndrome: clues from neuroimaging. In A. Klin, F. Volkmar and S. Sparrow (eds), *Asperger's Syndrome*. Guilford Press, New York, pp. 172–209.

Schweinberger, S.R. and Soukup, G.R. (1998). Asymmetric relationships among perceptions of facial identity, emotions, and facial speech. *J. Exp. Psychol. Hum. Percept. Perform*. **24**, 1748–1765.

Shields, J., Varley, R., Broks, P. and Simpson, A. (1996). Social cognition in developmental language disorders and high-level autism. *Dev. Med. Child Neurol*. **38**, 487–495.

Siegel, B., Vukicevic, J., Elliott, G.R. and Kraemer, H.C. *et al*. (1989). The use of signal detection theory to assess DSM-IIIR criteria for autistic disorder. *J. Am. Acad. Child Adolesc. Psychiat*. **28**, 542–548.

Sigman, M., Ruskin, E., Arbeile, S., Corona, R. *et al*. (1999). Continuity and change in the social competence of children with autism, Down's syndrome, and developmental delays. *Monogr. Soc. Res. Child Dev*. **64**, 1–114.

Simion, F., Valenza, E., Umilta, C. and Dalla Barbe, B. (1998). Preferential orienting to faces in newborns: a temporal–nasal asymmetry. *J. Exp. Psychol. Hum. Percept. Perform*. **24**, 1399–1405.

Sorce, J.F., Emde, R.N., Campos, J.J. and Klinnert, M.D. (1985). Maternal emotional signaling: its effects on the visual cliff behavior of 1 year-olds. *Dev. Psychol*. **21**, 195–200.

Stone, W.L. (1997). Autism in infancy and early childhood. In D.J. Cohen and F.R. Volkmar (eds), *Handbook of Autism and Pervasive Developmental Disorders*, 2nd edn. Wiley, New York, pp. 266–282.

Stone, W.L., Ousley, O.Y. and Littleford, C.L. (1997). Motor imitation in young children with autism: what's the object? *J. Abnorm. Child Psychol.* **25** (6), 475–485.

Swettenham, J., Baron-Cohen, S., Charman, T., Cox, A. *et al.* (1998). The frequency and distribution of spontaneous attention shifts between social and nonsocial stimuli in autistic, typically developing, and non-autistic developmentally delayed infants. *J. Child Psychol. Psychiat.* **39**, 747–753.

Szatmari, P., Tuff, L., Finlayson, A.J. and Bartolucci, G. (1990). Asperger's syndrome and autism: neurocognitive aspects. *J. Am. Acad. Child Psychiat.* **29**, 130–136.

Tanaka, J.W. and Gauthier, I. (1997). Expertise in object and face recognition. In R.L. Goldstone, P.G. Schyns and D.L. Medin (eds), *Psychology of Learning and Motivation*. Academic Press, San Diego, CA, pp. 83–125.

Tantam, D., Monaghan, L., Nicholson, H. and Stirling, J. (1989). Autistic children's ability to interpret faces: a research note. *J. Child Psychol. Psychiat.* **30**, 623–630.

Tomasello, M. (1995). Joint attention as social cognition. In: C. Moore and P. Dunham (eds.), *Joint Attention: Its Origins and Role in Development*. Erlbaum, Hillsdale, NJ, pp. 103–130.

Tomasello, M. and Farrar, J. (1986). Joint attention and early language. *Child Dev.* **57**, 1454–1463.

Tomasello, M., Kruger, A. and Ratner, H.H. (1993). Cultural learning. *Behav. Brain Sci.* **16**, 495–552.

Trad, P.V., Bernstein, D., Shapiro, T. and Hertzig, M. (1993). Assessing the relationship between affective responsivity and social interaction in children with pervasive developmental disorder. *J. Autism Dev. Disord.*, **23**, 361–377.

Ulvund, S. and Smith, L. (1996). The predictive validity of non-verbal communication skills in infants with perinatal hazards. *Infant Behav. Dev.* **19**, 441–449.

Valenza, E., Simion, F., Macchi Cassia, V. and Umilta, C. (1996). Face preference at birth. *J. Exp. Psychol. Hum. Percep. Perform.* **22**, 892–903.

Vecera, S.P. and Johnson, M.H. (1995). Gaze detection and the cortical processing of faces: Evidence from infants and adults. *Vis. Cognit.* **2**, 59–87.

Volkmar, F.R., Burack, J.A. and Cohen, D.J. (1990). Developmental vs. difference approaches to autism. In B. Hodapp, J. Burack and E. Zigler (eds), *Issues in the Developmental Approach to Mental Retardation*. Cambridge University Press, Cambridge, pp. 246–271.

Volkmar, F.R., Cohen, D.J., Bregman, J.D., Hooks, M.Y. and Stevenson, J.M. (1989). An examination of social typologies in autism. *J. Am. Acad. Child Adolesc. Psychiat.* **28**, 82–86.

Volkmar, F.R. and Cohen, D.J. (1985). The experience of infantile autism: a first-person account by Tony W. *J. Autism Dev. Disord.* **15**, 47–54.

Volkmar, F.R., Cohen, D.J. and Paul, R. (1986). An evaluation of DSM-III criteria for infantile autism. *J. Am. Acad. Child Psychiat.* **25**, 190–197.

Volkmar, F.R., Carter, A., Sparrow, S.S. and Cicchetti, D.V. (1993). Quantifying social development in autism. *J. Am. Acad. Child Adolesc. Psychiat.* **32**, 627–632.

Volkmar, F.R., Klin, A., Siegel, B., Statmari, P. *et al.* (1994). Field trial for autistic disorder in DSM-IV. *Am. J. Psychiat.* **151**, 1361–1367.

Volkmar, F.R., Sparrow, S.S., Goudreau, D., Cicchetti, D.V. *et al.* (1987). Social deficits in autism: an operational approach using the Vineland Adaptive Behavior Scales. *J. Am. Acad. Child Adolesc. Psychiat.* **26**, 156–161.

Volkmar, F.R., Carter, A., Grossmann, J. and Kin, A. (1997). Social development in autism. In D.J. Cohen and F.R. Volkmar (eds), *Handbook of Autism and Pervasive Developmental Disorders*. Wiley, New York, pp. 173–194.

Weeks, S.J. and Hobson, R.P. (1987). The salience of facial expression for autistic children. *J. Child Psychol. Psychiat.* **28**, 137–151.
Werner, E., Dawson, G., Osterling, J. and Dinno, N. (2000). Brief report: Recognition of autism spectrum disorder before one year of age: a retrospective study based on home videotapes. *J. Autism Dev. Disord.* **30**, 157–162.
Whalen, P.J., Rausch, S.L., Etcoff, N.L., McInerney, S.C. *et al.* (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *J. Neurosci.* **18**, 411–418.
Wicker, B., Michel, F., Henaff, M.A. and Decety, J. (1998). Brain regions involved in the perception of gaze: a PET study. *Neuroimage* **8**, 221–227.
Wing, L. and Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: epidemiology and classification. *J. Autism Dev. Disord.* **9**, 11–29.
Wing, L. (1981). Language, social, and cognitive impairments in autism and severe mental retardation. *J. Autism Dev. Disord.* **11**, 31–44.
WHO (World Health Organization) (1990). *International Classification of Diseases: Diagnostic Criteria for Research.* (draft version, 10th edn). World Health Organization, Geneva.

10

Do Children with ADHD not Need Their Frontal Lobes for Theory of Mind? A Review of Brain Imaging and Neuropsychological Studies

WINFRIED KAIN AND JOSEF PERNER
Department of Psychology, University of Salzburg, Austria

'Theory of mind' (ToM) is a label for our ability to impute mental states to ourselves and to others. Only by knowing what people desire, think, feel and intend can we understand social interactions and predict how people will behave. This ability is, therefore, at the heart of social cognition. In the last 20 years ToM has become an important topic of research efforts especially in developmental and clinical psychology (see e.g. Baron-Cohen, Tager-Flusberg and Cohen, 2000; Wellman, Cross and Watson 2001). In the last five years an increasing number of studies have set out to explore the neurophysiological basis of ToM in the brain. These studies provide the starting-point for this chapter. First we look at the current evidence for the neural substrate of ToM, which points to specific parts of the frontal lobes as serving a central function. Since the frontal lobes also appear essential for executive functions we look for evidence that executive tasks activate the same brain regions as are involved in ToM. We then touch on the developmental relationship between ToM and executive functions. Finally we look at the clinical group of children with Attention-Deficit Hyperactivity Disorder (ADHD), who are an interesting, yet under-researched group because of their well documented deficits in executive functions and social competence (see Barkley 1997).

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

THE BRAIN BASIS OF THEORY OF MIND

As animal models are inappropriate for investigating the neural substrate of mentalising (seen as a specifically human accomplishment) one can distinguish but two approaches: brain imaging (SPECT, PET, fMRI) during the solving of theory of mind (ToM) tasks, and studies of ToM performance by patients with specific brain lesions.

EVIDENCE FROM BRAIN IMAGING STUDIES

We were able to find 13 brain imaging studies of ToM (see Table 10.1). There are important differences between the studies in terms of resolution accuracy (1 SPECT, 7 PET, 5 fMRI), selection of ToM tasks, and type of participants (four studies used psychiatric patients).

The first study was undertaken by Baron-Cohen *et al.* (1994), who presented participants with two lists of words and participants had to raise their finger to indicate that they had heard a mind-related word for one list and a body-related word for the other list (control condition). Compared with the control condition, there was increased activity in the right orbito-frontal cortex [Brodmann area (BA)11] relative to decreased activity in the left frontal-polar region (BA10) during the mental state term recognition task. However, it should be noted that a region-of-interest approach was used and important brain regions were not explored that were shown relevant for ToM in later studies (e.g. BA8, BA9 and anterior cingulate).

All studies that employed written ToM stories (Fletcher *et al.* 1995; Happé *et al.* 1996; Gallagher *et al.* 2000; Vogeley *et al.* 2001) used the same careful design of comparing brain activation while reading a ToM story with activation in a control condition (reading a passage of unrelated sentences or physical events). All four of these studies agree that medial prefrontal cortex (BA8, extending into area BA9 and the anterior cingulate cortex) are uniquely activated while reading ToM stories. Some discrepancy among studies concerns hemispheric dominance: whereas in the study by Fletcher *et al.* (1995) and Happé *et al.* (1996) activation was seen predominantly on the left side, Vogeley *et al.* (2001) report activation of the anterior cingulate cortex in the right hemisphere, and Gallagher *et al.* (2000) mention no dominant side. Moreover, in the study by Happé *et al.* (1996) BA8 was not activated in patients with Asperger's syndrome (known to have subtle ToM deficits) but the adjacent BA9 and BA10 on the left side were.

Vogeley *et al.* (2001) also used additional 'self' stories. In the ToM + SELF condition ('self and other ascription stories') participants acted as one of the agents in the story. In the SELF condition ('self ascription story') participants read stories in which they themselves encountered an ambiguous situation and

Table 10.1. Brain imaging studies with theory of mind tasks

Study	Population	Tasks	Activation
Baron-Cohen *et al.* (1994) SPECT	12 males age 20–24	*E* Deciding if a word is a mental-related term or not *C* Deciding if a word is a body-related term or not	Increased activation in right orbito-frontal cortex (BA11) relative to decreased activity in the left frontal-polar region (BA10)
Fletcher *et al.* (1995) PET	Six right-handed males age 24–65, mean 38	*E* ToM stories *C* Unlinked sentences *C* Physical stories	Left medial frontal gyrus predominantly BA8, extending into posterior BA9 and the anterior cingulate cortex. Additional activation in the posterior cingulate cortex and inferior parietal lobe on the right (BA40)
Goel *et al.* (1995) PET	Five right-handed males, four right-handed females, mean 24.7	*E* ToM condition: how would someone with a background knowledge of Christopher Columbus infer the function of an artefact? *C* Baseline condition: visual perception and analysis, simple decision and motor response *C* Memory retrieval: presentation of modern familiar stimuli. Deciding if the stimuli are used for food preparation or personal care *C* Simple inference: presentation of unfamiliar stimuli. Deciding if the stimuli are used for food preparation or personal care	Selective activation of the left medial prefrontal cortex (BA9). Compared with simple inference condition: left posterior temporal lobe, left anterior temporal lobe and left medial frontal lobe
Happé *et al.* (1996) PET	Normal controls, five male Asperger's syndrome, mean 24 age 20–27	Same as Fletcher *et al.* (1995)	Normal controls: Left medial prefrontal cortex (BA8). Asperger's syndrome: no activation of BA8, but an adjacent, more ventral area of left medial prefrontal cortex (BA left9/10)

(*continued*)

Table 10.1. *Continued*

Study	Population	Tasks	Activation
Baron-Cohen *et al.* (1999) fMRI	six male normal controls, six female normal controls six persons with autism (four male, two female)	*E* Looking at photographs of eyes and deciding which of two words best described what the person is feeling or thinking *C* Looking at photographs of eyes and deciding whether they belong to a man or a woman	Fronto-temporal neocortical regions, comprising left dorsolateral (BA44/45/46), left medial prefrontal cortex (BA9); supplementary motor area and bilateral temporo-parietal regions (BA21/2/39/40). Non-neocortical areas: left amygdala, left hippocampal gyrus (BA27, 30), bilateral insulae and left striatum Autism: Less extensive activation of the frontal components and no activation of the amygdala at all
Gallagher *et al.* (2000) fMRI	Five right-handed males, one right-handed female, age 23–36, mean 30	Stories same as Fletcher *et al.* (1995) *E* ToM cartoons *C* Non-ToM cartoons *C* Jumbled pictures	ToM Stories: broader region of the medial prefrontal cortex extending into BA9 and closely associated with the anterior cingulate region of BA32 Cartoons: medial prefrontal cortex, restricted to BA8
Castelli *et al.* (2000) PET	Six right-handed males, age 20–31, mean 24.5	Watching computer-presented animations with two shapes and telling what was happening. Three animation types: *E* ToM *C* Goal-directed *C* Random action	More activity during ToM compared with random action: temporo-parietal junction (BA22/39), basal temporal region (BA37, 38), extrastriate cortex (BA19/18), medial prefrontal cortex (BA9) Differences more significant in right hemisphere, except for medial prefrontal cortex
Brunet *et al.* (2000) PET	Eight right-handed males, mean age 23.3	*E* Attributing intentions to the characters of a comic strip (AI) *C* Physical causality cartoons with characters PC–CH *C* Physical causality cartoons with objects PC–OB	Comparison AI vs PC–CH: right middle and medial prefrontal cortex including BA8/9; right inferior prefrontal cortex (BA47), right inferior temporal gyrus (BA20); left superior temporal gyrus (BA38); left cerebellum, bilateral anterior cingulate (BA24); and the middle temporal gyri (BA21)
Russell *et al.* (2000) MRI	Seven normal controls age 26–58, mean 40 Five right-handed patients with schizophrenia, age 26–57, mean 36	Same as Baron-Cohen *et al.* (1999)	Healthy subjects: left inferior frontal gyrus reaching into the insula (BA44/45/47), into the medial frontal lobe (BA45/9), in the left middle (BA21) and left superior temporal gyrus (BA22) Schizophrenia: Significantly less activation in the left middle/inferior frontal cortex (BA9/44/45)

Vogeley *et al.* (2001) fMRI	Eight right-handed males, age 25–36	Stories without SELF same as Fletcher *et al.* (1995) *E* Only ToM stories *E* Only SELF stories *E* ToM and SELF stories *C* Baseline condition: unrelated sentences *C* Stories without ToM/SELF, physical stories	Activation in ToM: predominantly in the right anterior cingulate and left temporo-polar cortex Activation in SELF: right temporoparietal junction and bilaterally anterior cingulate cortex + right premotor and motor cortex and precuneus ToM + SELF: right lateral prefrontal cortex
McCabe *et al.* (2001) fMRI	12 Right-handed subjects, comparison between seven co-operative and five non-cooperative subjects	Trust game Punish game Mutual advantage game Each with a human and computer counterpart	Cooperative subjects: activation differences between human and computer counterparts in the occipital lobe (BA17/18), parietal lobe (BA7), thalamus, middle frontal gyrus and the frontal pole (BA10) Uncooperative subjects: no activation differences in the medial prefrontal cortex between human and computer conditions
Castelli *et al.* (2002) PET	Able patients with autism or Asperger's syndrome, mean 33, 10 Control subjects, mean 25	Same as in Castelli *et al.* (2000)	Reduced activations in the autism group relative to control group: medial prefrontal cortex (BA9); Basal temporal area (BA20/38); temporo-parietal junction (BA22/40, 21)
Gallagher *et al.* 2002 PET	Nine males age 22–43, mean 32	Computer-based version of the playground game 'stone, paper, scissors' *E* Mentalising task: playing against the experimenter *C* Computer condition: playing against the computer — simple rule version *C* Computer condition: playing against the computer — responding randomly	Increased brain activity in anterior paracingulate cortex (BA32, 9/32)

Results mentioned refer to the selective activations in ToM tasks compared with the other tasks.
Abbreviations: *C*, control task; *E*, experimental task; BA, Brodmann area.

had to report their attitudes and perceptions. In comparison with the classic ToM stories, additional activation was found in the right temporo-parietal junction and the medial aspects of the superior parietal lobe (precuneus bilaterally).

Two studies also employed ToM cartoons. In the study by Gallagher *et al.* (2000), looking at a ToM cartoon was also associated with activation in the medial prefrontal cortex, but to a lesser extent and restricted to BA8. Additional activations were seen in the right middle frontal gyrus (BA6), the precuneus (BA7) and the cerebellum (left flocculus) when ToM cartoons were contrasted with jumbled pictures. However, these activations (with the exception of the medial prefrontal cortex) could be attributed to the results of the general requirements of interpreting meaningful cartoons, as they were also activated to a lesser extent by cartoons without ToM content.

Brunet *et al.* (2000) used stories consisting of three pictures and a fourth picture to be chosen by participants from an array of three response pictures. In the ToM cartoon condition, it was necessary to infer the intentions of the agent in order to select the correct response picture. In the two other story conditions (one involving characters and the others objects), only the comprehension of physical causality was required. As in other studies, significant rCBF increases in the ToM condition occurred in the right middle and medial prefrontal cortex (BA8 and BA9) and bilaterally in the anterior cingulate gyrus (BA24). Activation also occurred in the right inferior prefrontal cortex (BA47), in the anterior parts of the temporal lobes (BA20 right, BA21 bilaterally), in the left superior temporal gyrus (BA38) and in the left cerebellum. Because no control task comparable to jumbled pictures or meaningless cartoons was used, the additional activations could be attributable to the general processing of meaningful cartoons, as in the Gallagher study.

Castelli *et al.* (2000) developed an interesting non-verbal ToM task [similar to Heider and Simmel's (1944) animated cartoons]. They presented healthy subjects with abstract computer animations of geometric shapes. In one condition the shapes depicted 'random movements' and in another 'simple interaction'. In the ToM animation the complex interactions of the shapes evoked mental state attributions in the participants' descriptions of what was happening in the animations. Compared with the other two conditions, the ToM animations led to more bilateral activity in four main areas: medial prefrontal cortex (BA8/9), temporo-parietal junction (BA39), basal temporal region (BA37, BA34/38), and occipital cortex (BA17, BA18). The same paradigm was used with able adults with autism or Asperger's syndrome (Castelli *et al.* 2002). Compared with the normal control group, the autism group showed reduced activation in the medial prefrontal cortex (BA9), basal temporal area (BA20, BA38) and in the temporo-parietal junction (BA21, BA22/BA40).

Goel *et al.* (1995) presented a set of 150 stimuli of man-made old and modern artefacts and participants had to figure out how someone with a background knowledge of, say, Christopher Columbus would guess at the function of the

artefact. Compared to the three other conditions (see Table 10.1) the main difference was a selective activation of the left medial prefrontal cortex (BA9).

Two further studies used the 'eyes' task developed by Baron-Cohen *et al.* (1997) as an advanced ToM test and compared healthy subjects to patients with autism (Baron-Cohen *et al.* 1999) and to patients with schizophrenia (Russell *et al.* 2000). The eyes task activated two main components in the study by Baron-Cohen and colleagues: a set of fronto-temporal neocortical regions and a number of non-neocortical areas (see Table 10.1 for further description). The difference between healthy participants and patients with autism consisted of a lesser activation of the frontal components and complete lack of activation of the amygdala in the autistic group. Similarly, underactivation of the left prefrontal cortex (BA9/44/45) was also the main characteristic of patients with schizophrenia, in comparison to healthy comparison participants in the study by Russell *et al.* (2000).

Two recent studies (McCabe *et al.* 2001; Gallagher *et al.* 2002) examined mentalising ability while playing a game. In the study by McCabe *et al.* (2001) volunteers played three types of games (trust, punish and mutual advantage), each with a human and a computer counterpart. Afterwards, volunteers with the lowest and highest cooperation scores were compared. Significant activation differences between the human and computer conditions were seen in the medial prefrontal cortex, but only in the cooperative subjects.

In the study by Gallagher *et al.* (2002) volunteers played a computerised competitive game ('stone, paper, scissors') in three conditions (mentalising condition = playing against the experimenter; two comparison conditions = playing against a computer with a predetermined rule-based strategy or a random sequence). When comparing the mentalising and rule-solving conditions the only significant activation was seen in the anterior paracingulate cortex bilaterally (BA32, BA9/32).

Also an ERP (event-related potential using the EEG) study lends support to the role of the left frontal lobes in ToM. Sabbagh and Taylor (2000) found enhanced positivity over left frontal sites and a stronger negativity over left parietal sites, while processing short narratives requiring false belief understanding, in comparison to processing narratives about outdated, 'false' photos.

EVIDENCE FROM STUDIES WITH BRAIN-DAMAGED PATIENTS

Another line of evidence for the cerebral basis of ToM comes from neuropsychological studies with brain-damaged patients (see Table 10.2). From these studies one can infer which brain lesions produce ToM problems. However, there are — in comparison to the brain-imaging methodology — additional problems of interpretation, which need to be mentioned. There is greater variability of the effects of similar brain lesions because of the varying size of the lesions (and the number of disturbed connections to other regions).

Table 10.2. Neuropsychological studies with brain-damaged patients

Study	Population	Tasks	Problems
Siegal *et al.* (1996)	17 ps with unilateral RH damage 11 ps with unilateral LH damage	First-order true and false belief with 11 RH and six LH First-order true and false belief (look first) with six RH and five LH	Nearly all RH and LH passed false belief—look first tasks Significant more problems with the other false belief tasks in RH than in LH
Winner *et al.* (1998)	13 ps with unilateral RH damage 20 controls	First-order true belief Second-order belief Distinguishing jokes from lies	More first-order true belief errors in RH (but no significant difference) Significant difference in attributing second-order beliefs No difference in second-order expectation question
Stone *et al.* (1998)	Five ps with bilateral OFC damage Five ps with unilateral left dorso-lateral pre-frontal cortex damage Five controls age-matched	First-order belief Second-order belief Comprehension of social *faux pas*	No differences on first-order and second-order beliefs between groups Comprehension of *faux pas* difficult for most OFC patients, but not for the other two groups
Happé *et al.* (1999)	Study 1: 14 ps with RH damage; 19 healthy elderly controls Study 2: five ps with LH damage	Advanced ToM stories Non-mental stories ToM cartoons Non-mental cartoons Cartoon pairs	Study 1: no difference to controls on non-mental stories, but on ToM stories; great difference on ToM cartoons, little difference on non-mental cartoons, compared with controls; difference on ToM cartoon pairs Study 2: no difficulty on ToM stories or cartoons compared with non-mental tasks
Corcoran (2000)	Five epileptic ps RF or FT foci Three epileptic ps with LF or FT foci Three epileptic ps with BF foci or damage 23 controls	Hinting task: inferring the intention of a speaker	Significant difference between RF/FT and controls in appreciation of hints

Channon and Crawford (2000)	Six ps with unilateral LF lesions 13 ps with unilateral RF lesions Four ps with LP lesions Eight ps with RP lesions 60 Controls	Brief written vignettes on sarcasm, pretence, misunderstanding, lie, white lie, threat and dare Explanation why someone has said this	Left anterior group impaired compared with all other groups
Rowe *et al.* (2001)	15 ps with RF lesions 16 ps with LF lesions 31 controls	Six first-order false belief Six second-order false belief	Impairment compared with control, but no difference between RF and LF on first-order and second-order false belief No effect of size of lesions
Stuss *et al.* (2001)	Four ps with RF lesions Eight ps with LF lesions Seven ps with BF lesions Five ps with RNF lesions Eight ps with LNF lesions 14 controls	One Baseline task Two visual perspective tasks; direct inference task and transfer inference task One deception task	No group differences for baseline and direct inference task More errors on transfer inference task in frontal patients (most in ps with RF) More errors in ps with BF, no difference to controls for ps with RF or LF
Surian and Siegal (2001)	Ps with RH damage Ps with LH damage	ToM tasks presented verbally and with visual aids	Good performance of RH and LH with visual aids Impairment in RH patients when presented only verbally

Abbreviations: ps, patients; RH, right hemisphere; LH, left hemisphere; RF, right frontal; LF, left frontal; BF, bilateral frontal; OFC, orbito-frontal cortex; FT, fronto-temporal; RNF, right non-frontal; LNF, left non-frontal.

Therefore, the conclusions from these studies can only be interpreted with caution. One can generally distinguish between studies in which patients with varying and multiple unilateral lesions were selected (Siegal, Carrington and Radel 1996; Winner *et al.* 1998; Happé, Brownell and Winner 1999; Surian and Siegal 2001) and studies with patients with circumscribed unilateral lesions, e.g. anterior or posterior area (Stone, Baron-Cohen and Knight 1998; Channon and Crawford 2000; Corcoran, 2000; Rowe *et al.* 2001; Stuss *et al.* 2001). This is an important difference, because the latter type of studies allows greater precision in inferring the brain areas involved in ToM. In addition, there are single case studies with patients with focal brain damage (Bach *et al.* 1998, 2000; Blair and Cipolotti 2000; Fine, Lumsden and Blair 2001; Happé, Malhi and Checkley 2001; Lough, Gregory and Hodges 2001).

The general result from the multiple lesions studies is that patients with right hemisphere damage (RHD) have problems with ToM, but to different degrees. In the study by Siegal *et al.* (1996), RHD patients had significantly more problems with first-order false-belief tasks than patients with left hemisphere damage (LHD). This was not the case with first-order false-belief tasks inserting the adjunct 'look first'. This is an enormous deficit, bearing in mind that many children as young as 4 years pass this task (Perner, Leekam and Wimmer 1987; Wellman, Cross and Watson 2001). In another study, Surian and Siegal (2001) also employed first-order false-belief tasks with either implicit or explicit ('look first') questions, but this time the tasks were also presented with visual aids (pointing to a blue or white card). There was no significant difference between RHD and LHD patients. Surian and Siegal attribute these results to impaired visuospatial buffers and working memory deficits, rather than to a fundamental ToM deficit.

In a study by Winner *et al.* (1998), there was no significant difference between RHD patients and healthy controls on first-order true-belief tasks, in contrast to significant differences on second-order false-belief tasks (e.g. distinguishing lies from jokes). In a study by Happé, Brownell and Winner (1999), RHD patients also performed worse on advanced ToM tests and on ToM cartoons, compared with elderly controls. A sample of five LHD patients did not show any problems on these tasks. Finally, Corcoran (2000) reports data on epileptic patients with right frontal or fronto-temporal foci (RF/FT), left frontal or fronto-temporal foci (LF/FT) or bilateral frontal foci (BF). Using the Hinting Task (vignettes: one character gives a heavy hint from which another character has to infer his intention), attenuated performance was found in the RF/FT group compared with the other two groups.

A somewhat different picture emerges when one looks at the studies on patients with circumscribed lesions. Here, the difference between right and left hemispheres disappears in favour of the importance of the right *and* left frontal lobes. Two studies focused only on patients with frontal lesions. In the first study, Stone, Baron-Cohen and Knight (1998) compared: five patients with damage to

the left lateral frontal cortex, including dorsal regions (DFC group); five patients with bilateral damage to the orbitofrontal cortex (OFC group); and five age-matched controls. There was no difference between the groups on first-order and second-order false-belief tasks when memory load was minimal (no memory load condition). Having to remember the stories was only a problem for the DFC group. On the other hand, the DFC group correctly solved another advanced ToM test ('*faux pas* task'), whereas almost all OFC patients made errors on this task. Another study employed first-order and second-order false-belief tests, comparing patients having exclusively right-sided and exclusively left-sided frontal lobe lesions with healthy controls (Rowe *et al*. 2001). There was no difference between the two patient groups on the false-belief tasks but they were both significantly impaired compared with the control group.

Two other studies compared the effects of anterior and posterior lesions on ToM. In the study by Channon and Crawford (2000), four patient groups (with left anterior, right anterior, left posterior and right posterior lesions) and a healthy control group received brief written vignettes and had to explain the main character's utterance or actions. Only the group with left anterior lesions had significant problems with these vignettes, whereas the other three groups did not differ significantly from the control group. Stuss, Gallup and Alexander (2001) compared five patient groups (right frontal, left frontal, bifrontal, right non-frontal and left non-frontal) with a healthy control group. They used two different visual perspective tasks of varying inference complexity about the visual experience of others (direct inference and transfer inference condition) and a deception task requiring first-order attributions. Patients with frontal lesions committed more errors in the transfer inference condition than the posterior and healthy groups, which did not differ. As to frontal laterality, the results are not very clear. Although more problems were found in the right than the left frontal group, this effect seems to be attributable to the bifrontal group, who had the greatest problems. On the deception task, only the bifrontal and the two posterior groups had problems.

Overall, the following picture emerges from studies on brain-damaged patients. If patients with generalised unilateral damage (including diverse frontal, temporal and parietal lesions) are selected, clear problems in advanced ToM tasks are found in patients with right hemisphere damage in all studies compared to patients with left hemisphere damage and/or controls (Winner *et al*. 1998; Happé, Brownell and Winner 1999). Problems with first-order belief tasks were only found in the study by Siegal *et al*. (1996), but this seems attributable to general information-processing deficits, as these problems disappeared with explicit task questions ('look first') and visual aids (Surian and Siegal 2001). On the other hand, the results from studies on patients with circumscribed lesions in a particular area (left vs. right frontal, anterior vs. posterior regions) point to the dominant role of the frontal lobes in general. Concerning laterality in these studies, three of the four studies found deficits in patients irrespective of laterality

(Stone, Baron-Cohen and Knight 1998; Stuss, Gallup and Alexander 2001; Rowe *et al*. 2001), whereas one study (Channon and Crawford 2000) found ToM deficits only in patients with left anterior lesions (in contrast to the studies comparing patients with generalised unilateral damage!).

WHICH OTHER TASKS ACTIVATE 'ToM AREAS'?

There is substantial agreement in existing brain-imaging studies that mentalising is associated with particular circumscribed brain regions uniquely activated in ToM tasks: medial prefrontal cortex, comprising BA8 and BA9. More recent reviews, however, see relevant activations restricted to the anterior paracingulate cortex, BA 9/32 (Frith and Frith 2001; Gallagher and Frith in press). In five of 13 studies, activations in these regions were seen in the left medial prefrontal cortex (Fletcher *et al*. 1995; Goel *et al*. 1995; Happé *et al*. 1996; Baron-Cohen *et al*. 1999; Russell *et al*. 2000). Five studies refer generally to the medial prefrontal cortex or anterior paracingulate sulcus with no dominant hemisphere mentioned (Castelli *et al*. 2000, 2002; Gallagher *et al*. 2000, 2002; McCabe *et al*. 2001). Finally, three studies report activations predominantly on the right side, with different loci mentioned: right orbitofrontal (Baron-Cohen *et al*. 1994), right middle and medial (Brunet *et al*. 2000), and right lateral PFC (Vogeley *et al*. 2001). Activations are also frequently reported in the anterior cingulate cortex (without differentiating it from the anterior paracingulate cortex), superior temporal sulcus and the temporal poles bilaterally.

These findings raise the interesting question of what other cognitive tasks activate medial prefrontal cortex (BA8 and BA9) or the anterior paracingulate cortex. Not only activation in ToM tasks but also activation in self-referential mental activity requires the understanding of other minds. As Gallagher and Frith (in press) note, the paracingulate cortex was also activated in different kinds of self-monitoring, such as visual self-recognition, memory for autobiographical events, verbal self-monitoring, self-generated thoughts, externally-produced tickling and perception of pain. Recently, Johnson *et al*. (2002) also found activations in the medial prefrontal cortex while volunteers reflected on their abilities, traits and attitudes.

Additionally, Lane (2000) postulates that the dorsal anterior cingulate cortex plays a dominant role in the direct experience of an emotion (phenomenal awareness of emotion), whereas the rostral anterior cingulate cortex (BA 32) and medial prefrontal cortex (BA 9) are activated during selective attention to subjective emotional responses (reflective awareness of emotion). This is supported by a recent meta-analysis of PET and fMRI studies on emotion (Phan *et al*. 2002), in which the medial prefrontal cortex was strongly involved in the cognitive aspects (attention to and appraisal of emotion) of emotional

processing. Interestingly, Berthoz *et al.* (2002) corroborate this finding in a clinical sample. Comparing men with and without alexithymia (poor expressiveness of emotional states), they found significant activation differences in the medial prefrontal cortex/paracingulate and anterior cingulate cortex in response to high-arousal stimuli of negative or positive valence.

In a review of 275 PET and fMRI studies, Cabeza and Nyberg (2000) report that BA8 and BA9 are typically activated (about equally often left and right) in problem-solving tasks such as the Wisconsin Card Sorting Test or the Tower of London task. BA9 was additionally involved in sustained attention, working memory, fluency tasks and episodic memory retrieval. Most of these tasks, especially those involving executive functions such as working memory or problem solving, are mainly activated in *dorsolateral* part of BA9 or in the dorsolateral prefrontal cortex involving BA9/46 and BA10 (see also Dagher *et al.* 1999; Rowe *et al.* 2001; for review on working memory, see Colette and Van der Linden 2002). In contrast, most activations in ToM tasks lie in the *medial* part of BA8 and BA9 or, more precisely, in the anterior paracingulate cortex (Frith and Frith 2001). Only a few studies report activations in 'ToM areas' for classical executive function tasks. These tasks are especially tasks which require the inhibiton of a prepotent response like the Stroop or the Go/No-Go task.

One study by de Zubicaray *et al.* (2000) used a verbal task requiring the inhibition of prepotent responses. In this study, subjects had to nominate in the control task the appropriate general superordinate category to which a word belonged (e.g. broccoli–vegetable), whereas in the inhibition task they had to nominate a general superordinate category to which a word did not belong (e.g. nose–animal). Looking at the Talairach values, increased activation in the inhibition task was seen in 'ToM areas' although the authors labelled it left *dorsolateral* prefrontal cortex (BA9). Using the classical Stroop interference tasks, Audenaert *et al.* (2001) found significant activation in the left medial prefrontal cortex (BA8).

Two studies with Go/No-Go tasks found activations of the medial prefrontal cortex (Liddle, Kiehl and Smith 2001; Menon *et al.* 2001). Analysing the activations in Go/No-Go trials, both studies come to the same conclusion, that the activation of the medial prefrontal cortex is not related to the response inhibition *per se* but to error processing, decision formation and monitoring.

Although we are still far away from understanding the exact role that medial BA8, BA9 and the anterior paracingulate cortex play in cognition (because of co-activation of other brain regions and the many different cerebral networks involved in different cognitive tasks), current evidence suggests the following picture. Besides its prominent role in attributing mental states to others, these regions are also activated in self-referential mental activities such as self-monitoring, self-reflection and attention/awareness of emotional processing.

With regard to so-called executive function tasks there are only a few studies which report activations in the ToM areas. Primarily these are tasks which require responding to incongruent stimuli and hence conflict and error monitoring like Stroop tasks or some Go/No-Go tasks. As Bush, Luu and Posner (2000) note, processing conflict or competition is often attributed to the anterior cingulate cortex of which the most anterior part is the paracingulate cortex. But Bush, Luu and Posner (2000) refer to the cognitive divison of the anterior cingulate cortex, which is primarily activated in cognitive Stroop-like tasks. As Frith and Frith (2001) remark, this cognitive division is generally posterior to the ToM areas, as distinguished from the affective division of the anterior cingulate cortex, which is also activated during ToM tasks. In contrast, other executive tasks, such as working memory or problem solving/planning (Wisconsin Card Sorting Task or Tower of London task) are not activated in the ToM areas but are primarily activated in the dorsolateral prefrontal cortex.

THE DEVELOPMENTAL RELATIONSHIP BETWEEN ToM AND EXECUTIVE FUNCTIONS

The link between executive functions and ToM is currently under hot debate in developmental psychology (Perner and Lang 1999; Moses 2001). So far, two kinds of executive functions have been specifically explored for their association with theory of mind development around 3–5 years: working memory and inhibitory control. In all previous studies, ToM tasks correlated significantly with working memory (Davis and Pratt 1996; Gordon and Olson 1998; Keenan 1998) and inhibitory control (Frye, Zelazo and Palfai 1995; Carlson and Moses 2001). There are three studies that used both working memory and inhibitory control tasks. Hughes (1998) found that both working memory and inhibitory control correlated significantly with all three ToM measures used (false belief prediction, false belief explanation, deception). However, once age and verbal/non-verbal intelligence was controlled, only the relation between deception and working memory and between deception and inhibition remained. Similarly, in the Carlson, Moses and Breton (2002) study, there was a significant link between working memory and inhibitory control (requiring conflict, not delay) tasks and two types of ToM measures (appearance reality and false belief prediction). But when controlling for age and overall IQ, the association between working memory and the two ToM measures disappeared, whereas it remained significant for inhibitory control and false belief prediction. They point to a specific role of inhibition above working memory contributing to ToM performance. White and Keenan (unpublished) compared working memory and inhibitory control with four ToM measures (false belief, representational change, appearance–reality and

deception). They generally found high intercorrelations between all measures. Controlling for age and language ability again reduced most correlations, leaving the significant relation between working memory or inhibitory control and false belief, as well as inhibitory control and deception.

Overall, these studies indicate that working memory and inhibitory control are associated with ToM performance in an important way, although the specific causal relations among these factors, and also with general verbal and non-verbal abilities, are far from clear.

ADHD, EXECUTIVE FUNCTIONS AND ToM

WHY IS RESEARCH OF ToM IN ADHD TIMELY?

Apart from the patients with brain damage, discussed above, two other clinical disorders exhibit great problems with imputing mental states to other people: autism spectrum disorders and schizophrenia (see Baron-Cohen, Tager-Flusberg and Cohen 2000; Blackwood *et al*. 2001; Corcoran 2001). Children and adults with these disorders have also great problems with a variety of executive functions (see Russell 1997), so that one is tempted to conclude that impairment in theory of mind is necessarily linked to impairment in executive functions. This is also underlined by the finding from normal development that progress in ToM around the age of 4 years relates specifically to progress in executive control at this age (for review, see Perner and Lang 1999). These findings make children and adults with attention-deficit-hyperactivity disorder (ADHD) a most interesting group to investigate, because they are known for their executive problems (see Barkley 1997) but have, until recently, rarely been investigated for their ToM ability.

In order to investigate this issue, we look first at neurophysiological evidence to see whether the medial prefrontal cortex or anterior cingulate cortex are also implicated in the aetiology of ADHD. We then briefly review neuropsychological evidence on executive deficiencies in ADHD.

Evidence from Neuroimaging Studies

One can generally distinguish two types of neuroimaging studies on patients with ADHD. Structural imaging explores general anatomical differences (e.g. volume, asymmetry) in the brain, and functional imaging explores the brain activity in different regions.

In their reviews of structural imaging studies, Castellanos (2001) and Giedd *et al*. (2001) list several brain regions that, in most studies, have been found to be of reduced volume in patients with ADHD (almost exclusively male), compared with controls. These regions include the right prefrontal brain,

anterior part of the corpus callosum, caudate nucleus, globus pallidus and a subregion of the cerebellar vermis. In a recent study by Castellanos *et al.* (2001) with 50 ADHD girls and 50 controls, reduced volumes were only found in the left caudate and posterior–inferior vermis after controlling for total cerebral volume and vocabulary. It remains an open question how much structural abnormalities in ADHD are influenced by gender or general intellectual differences (e.g. vocabulary). Moreover, critics point out inconsistencies with earlier studies concerning the brain regions affected, and criticise the small number of independent studies and statistical methods employed (see Baumeister and Hawkins 2001).

Although there is increasing evidence that particular brain regions are structurally different in ADHD patients compared to controls, the question remains how and why these differences are related to the core symptoms and executive deficits of ADHD. The only studies that have tried to explore the links between structural brain characteristics and executive functioning in ADHD were undertaken by Casey *et al.* (1997) and Semrud-Clikeman *et al.* (2000). Casey and colleagues compared the performance of ADHD and control children in three different response inhibition tasks with the size of particular brain regions. Two important findings emerged: first, there was a significant correlation between response inhibition and volumetric measures of the prefrontal cortex (defined as 'all brain matter in front of the anterior-most point of the corpus callosum'), caudate nucleus and globus pallidus, predominantly in the right hemisphere. Second, whereas the volumetric measures of the caudate and globus pallidus correlated with the performance in the control and inhibition trials, the size of the right prefrontal cortex was specifically associated with inhibition trials. This indicates a prominent role of this region in ADHD children for suppressing prepotent answers.

Using a regions-of-interest (ROI) approach, Semrud-Clikeman *et al.* (2000) reported a relationship between caudate asymmetry and poorer performance on the Stroop test, as well as more frequent failures to maintain set in the Wisconsin Card Sorting Test in both ADHD and control children. This relationship was more pronounced in ADHD children, although a statistical analysis between the ADHD and control group was not undertaken because of the small sample size. A relationship between anterior–superior regions and these tasks was not reported.

Functional imaging studies in ADHD can be grouped into three categories: studies comparing general metabolic activity in the brain of ADHD patients and controls, studies looking at metabolic changes during the performance of executive tasks, and studies exploring the effects of stimulants used as the primary medical treatment for ADHD.

The two research groups of Lou and of Zametkin conducted the first studies on cerebral blood flow in ADHD. Lou, Henriksen and Bruhn (1984) and Lou *et al.* (1989), using xenon inhalation and computed tomography, found

decreased blood flow in prefrontal regions and the striatum. Zametkin *et al.* (1990; Ernst *et al.* 1994, 1998), using PET, also reported reduced metabolic activity, especially in the left prefrontal cortex in adults and female adolescents with ADHD, although this could not be replicated in male adolescents with ADHD (Zametkin *et al.* 1993) and a larger sample of female adolescents with ADHD (Ernst *et al.* 1997). Overall, the studies of the research group of Zametkin and Ernst indicate that reduced metabolic activity is more pronounced in adults than in children and that it is influenced by both sexual maturation and intellectual differences (see Ernst *et al.* 1997). More recently, Spalletta *et al.* (2001) found reduced regional cerebral blood flow in the left dorsolateral prefrontal cortex in ADHD children.

Studies that explored metabolic activity during the performance of executive tasks have used different forms of inhibition tasks (Vaidya *et al.* 1998; Bush *et al.* 1999; Rubia *et al.* 1999, 2001; Langleben *et al.* 2001) and working memory tasks (Schweitzer *et al.* 2000). Vaidya *et al.* (1998) used two versions of a Go/No-Go task with ADHD children and controls and found greater bilateral activation of the frontal cortex on the response-controlled version (Go and No-Go blocks were equated in the number of key presses but differed in the number of trials and rate of stimulus presentation) and reduced striatal activation in the stimulus-controlled version (Go and No-Go blocks were equated for the rate of presentation and number of trials but differed in the number of key presses) in ADHD children compared to controls. The authors attributed the surprising result of *hyper*metabolism of the frontal regions (in contrast to *hypo*metabolism found in past studies) to greater inhibitory effort in ADHD children. Using the same task as Vaidya *et al.* (1998), Langleben *et al.* (2001) found a rCBF decrease in the right prefrontal cortex (BA9, BA44, BA46) relative to the left in ADHD children with severe or moderate hyperactivity, whereas this was not the case in ADHD children with low hyperactivity.

In another study by Rubia *et al.* (1999), using a Stop and Delay task, reduced activation was seen during the stop task in ADHD adolescents in the right mesial frontal cortex (BA8/32) at the border with the anterior cingulate cortex, right inferior and medioinferior frontal lobe (BA45 and BA9/45) and left caudate nucleus, as compared to controls. In a similar vein, Rubia *et al.* (2001) found significantly more activation in the right medial and inferior frontal cortex (BA9/45), the right mesial frontal cortex (BA8/32) and the left caudate nucleus in a control group compared to ADHD adolescents. Bush *et al.* (1999) used a counting Stroop task (control condition, reporting via key-press the number of animal words from 1 to 4; interference condition, words consisted of number words). Employing a region-of-interest approach focusing on the anterior cingulate cortex, they found greater activation in a network consisting of the anterior cingulate cortex (ACcd, cognitive division), left lateral prefrontal cortex (BA9) and superior parietal cortex (BA7) in control subjects. In contrast, adults with ADHD did not specifically activate this network.

The discrepant findings concerning lateralisation seem to be attributable to the tasks used. Vaidya *et al.* (1998) used a Go/No-Go task with letters and found bilateral frontal activation. The counting Stroop in the study by Bush *et al.* (1999) also requires verbal processing, leading to greater activation on the left side. In contrast, the Stop Task in the studies by Rubia *et al.* (1999, 2001) involved visual objects (e.g. pressing the button when an airplane appeared alone and not pressing when it was followed by a bomb) and so greater activation of the right frontal hemisphere seems logical.

Using a working memory task (Paced Auditory Serial Addition Task), Schweitzer *et al.* (2000) found rCBF increases in the anterior cingulate and medial frontal regions (BA32/10) and decreases in the left middle frontal regions (BA9) in controls, whereas ADHD adults showed decreases in the left middle temporal lobe (BA21) and increases in the right lenticulate, left parahippocampal gyrus (BA35/36) and bilaterally in the cerebellum.

Two neuroimaging studies explored the metabolic activity in children with ADHD after receiving stimulants. In general, stimulants have a 75–95% positive effect on the core symptoms in patients with ADHD (Solanto, Arnsten and Castellanos 2001). Mehta *et al.* (2000) showed that in normal human volunteers methylphenidate, as compared to a placebo, caused improvements on a self-ordered spatial working memory task and coincided with reduced rCBF in the left dorsolateral prefrontal cortex and the left posterior parietal cortex. The reduction of activation in these regions was interpreted by the authors as a result of increased efficiency in performing the task. Vaidya *et al.* (1998) found increased frontal activation during inhibition tasks in ADHD and control children after receiving methylphenidate. In control children methylphenidate caused frontal activation only in one particular inhibition task and decreased striatal activation.

What is the overall picture that emerges from neuroimaging studies of the brain mechanisms involved in ADHD? Before answering this question, it must be stated that it is premature to speak of a clear picture, simply because there are too few studies and those few employ too many different methodologies, which leaves us with more questions than answers. Of particular concern are differences in selected populations (children vs. adults; male vs. females) of manifestations of ADHD (predominantly inattentive vs. combined vs. predominantly hyperactive–impulsive), of co-morbidity (e.g. conduct disorder, learning disabilities), of general verbal and non-verbal intellectual abilities, and differences in executive tasks used (as well as a lack of replication studies). Despite these qualifications there is, nevertheless, strong evidence for the involvement of a fronto-striatal network in producing the core symptoms of ADHD (see also Solanto *et al.* 2001).

Looking at the link between ToM and ADHD, we have to apply to the ADHD studies the same differentiation between dorsolateral cortex and medial prefrontal cortex including anterior cingulate and paracingulate cortex.

Applying this differentiation is hampered by the fact that most brain imaging studies with ADHD do not specify the Talairach values or even Brodmann areas. From the functional imaging studies reviewed, we found five studies (Vaidya *et al.* 1998; Bush *et al.* 1999; Rubia *et al.* 1999, 2001; Schweitzer *et al.* 2000) which report activations in or near areas that are also found in ToM tasks. The results of Bush *et al.* (1999) leave it open whether in ADHD the cognitive division of the anterior cingulate cortex is more involved than the emotional division (which lies closer to the ToM areas). Additional studies refer to the important role of the dorsolateral prefrontal cortex (Mehta *et al.* 2000; Spalletta *et al.* 2001).

Evidence from Neuropsychological Studies

Another, indirect line of evidence for possible links between ToM and ADHD comes from classical neuropsychological studies. There are already substantial reviews of evidence that children with ADHD have significant problems with executive tasks (Barkley, Grodzinsky and DuPaul 1992; Pennington and Ozonoff 1996; Barkley 1997; Oosterlaan, Logan and Sergeant 1998; Nigg 2001). Pennington and Ozonoff (1996) reviewed 18 studies and found the greatest executive function (EF) deficits of children with ADHD in the Tower of Hanoi test (mean effect size of $d = 1.08$), in the error scores of the Matching Familiar Figures Test ($d = 0.87$), in various motor inhibition tasks ($d = 0.85$) and in the Time measure of the Trailmaking Test Part B ($d = 0.75$). It should be noted that some EF measures, like different working memory tasks, had not been included in their analysis. There is also more recent evidence of deficits in working memory, planning ability and inhibition in children with ADHD that has appeared since these reviews were published.

ADHD and Working Memory

In his review of working memory deficits in ADHD, Barkley (1997) concludes that there is strong evidence for deficits of verbal working memory in ADHD but less impressive evidence for deficits of non-verbal working memory (partly because of the scarcity of relevant studies). In the reviews by Nigg (2001), Oosterlaan, Logan and Sergeant (1998), and Pennington and Ozonoff (1996), measures of working memory were not explicitly considered.

Most recent studies (see Table 10.3) show that working memory deficits in ADHD are heavily influenced by general factors, such as verbal ability and intelligence. Studies by Cohen *et al.* (2000) and Willcutt *et al.* (2001) underscore the importance of verbal ability. Only ADHD children with language impairment (Cohen *et al.* 2000) and with co-morbid reading disability (Willcutt *et al.* 2001) exhibited deficits in verbal and non-verbal working memory, whereas this was not the case for pure ADHD children. However, although

Table 10.3. Overview of studies (1997–2001) on working memory and ADHD

Study	Population	Tasks	Results
Roodenrys *et al.* (2001)	16 ADHD and RD 16 RD 16 Controls	Verbal working memory: Memory span task Children's Paced Auditory Serial Addition Task (CHIPASAT) Memory updating task	ADHD/RD + RD < control ADHD/RD < RD < control ADHD/RD < control
Willcutt *et al.* (2001)	35 ADHD children 26 ADHD and RD 53 RD children 84 Controls	Verbal working memory: Sentence span Counting span	ADHD + RD, RD < control, pure ADHD
Kuntsi *et al.* (2001)	51 ADHD children 118/119 controls	Verbal working memory: Sentence span Counting span Non-verbal working memory Delayed response alternation task (DRA), pre- and postinstruction version	Only sentence span: ADHD < control, n.s. when IQ was controlled DRA: ADHD < control n.s. when IQ was controlled
Murphy *et al.* (2001)	105 Young adults with ADHD 64 Controls	Verbal working memory: Digit span from WAIS-III (but combined) Non-verbal working memory: Simon game	ADHD < control, no difference when IQ was controlled ADHD < control
Barnett *et al.* (2001)	27 Unmed. ADHD children 21 Med. children 26 Controls	Non-verbal working memory: Computerised tests of spatial working memory	Unmed. ADHD < controls

Cohen *et al.* (2000)	69 ADHD children 36 ADHD and LI 30 OPD and LI 31 OPD	Verbal working memory: Sentence span Counting span Non-verbal working memory: Recall of increasing numbers of spatial stimuli	ADHD + LI, OPD + LI < ADHD, OPD ADHD + LI, OPD + LI < ADHD, OPD
Kempton *et al.* (1999)	15 Unmed. ADHD children 15 Med. ADHD children 15 Controls	Non-verbal working memory: Self-ordered searching task	Unmed. ADHD < control on between-search errors
Milch-Reich *et al.* (1999)	38 ADHD children 41 Controls	Verbal working memory: On-line representation and organisation of picture stories	ADHD < control
Øie *et al.* (1999)	20 ADHD adolescents 19 Adolescents with schizophrenia 30 Controls	Verbal working memory: Digit span backwards Digit span distractibility task	ADHD = schizophrenia + Controls ADHD < control
Karatekin and Asarnow (1998)	31 ADHD adolescents 13 Adolescents with schizophrenia 27 Controls	Verbal working memory: Digit span backwards Non-verbal working memory: Delayed recall on the Dot Test	ADHD = control ($p = 0.072$) ADHD, schizophrenia < controls

Abbreviations: n.s, not significant; (Un)med., (un)medicated; ADHD, attention-deficit/hyperactivity disorder; RD, reading disability; LI, language impairment; OPD, other psychiatric disorders.

Roodenrys, Koloski and Grainger (2001) did not recruit pure ADHD children in their study, they found greater impairment in verbal working memory in ADHD children with reading disability than in children with reading disability alone.

Studies that controlled for IQ (Kuntsi, Oosterlaan and Stevenson, 2001; Murphy, Barkley and Bush, 2001) found that verbal working memory deficits disappeared when IQ was controlled. The same picture emerged also for non-verbal working memory in the study by Kuntsi, Osterlaan and Stevenson (2001), in contrast to the study by Murphy, Barkley and Bush (2001), in which significant differences remained.

Øie, Sundet and Rund (1999) compared ADHD adolescents with adolescents with schizophrenia and with normal controls on several memory measures. Concerning working memory, they found no significant differences on the digit span backward between groups, although the ADHD children performed worst. The only significant deficits in ADHD adolescents were found in a digit span distractibility task. Using the backward digit span task, Karatekin and Asarnow (1998) also found a non-significant difference ($p = 0.07$) between ADHD adolescents and controls, whereas there was a significant difference on two non-verbal working memory tasks. Two further studies explored the effects of stimulants on non-verbal working memory (Kempton *et al.* 1999; Barnett *et al.* 2001), with the same result, that non-medicated ADHD children had significant deficits compared to medicated ADHD children and controls.

An additional study is worth mentioning. Milch-Reich *et al.* (1999) presented ADHD and control children with picture stories and then recorded their verbal description of each picture at presentation and their later free recall of the whole story. Although this was not a classical working memory task, the procedure required working memory for integrating new information in an 'on-line representation' and for recall of the whole story. In comparison with control children, ADHD children generated fewer between-picture links and their on-line integration was inferior and less organised.

ADHD and Planning and Set-shifting Tasks

The Tower of Hanoi and Tower of London tasks are considered classic planning tasks that get to the heart of planning disorders (Lezak 1995). In the review by Pennington and Ozonoff (1996), it was the measure that yielded the strongest difference between ADHD children and controls. Since then, six further studies have compared ADHD children with controls on Tower tasks. Two studies with the Tower of Hanoi (Aman, Roberts and Pennington 1998; Klorman *et al.* 1999) confirm the planning deficit of ADHD children. However, clarification is required as to whether this deficit exists for all ADHD children or only for certain subtypes. Klorman *et al.* (1999) found no

differences between ADHD children of the predominantly inattentive type and non-ADHD children, whereas children with ADHD combined type had significantly fewer solutions and more rule violations than non-ADHD children. Three further studies with the Tower of London task complement this picture. Culbertson and Zillmer (1998) found significant differences between ADHD children and controls on three measures of their Tower of London study: general solutions, rule violations, and time violations. Korkman, Kirk and Kemp (1998) also report significant differences between ADHD children and controls in their validation of the NEPSY, and Kempton *et al*. (1999) found comparable performance in medicated ADHD children and controls, whereas unmedicated ADHD children had significant problems. However, in contrast to these studies, Houghton *et al*. (1999) did not find any differences between ADHD children and controls.

Another classical measure of executive function is the Wisconsin Card Sorting Test (WCST), which requires 'abstract behaviour' and set-shifting (Lezak 1995). In their review Barkley, Grodzinsky and DuPaul (1992) found significant differences on the WCST in eight of 13 studies. An additional eight studies were reviewed by Barkley (1997), finding differences in five of these. Of six more recent studies (Table 10.4), all show clear differences. In four of them, ADHD children had significantly more problems (especially in the category of perseverative errors) than controls (Pineda *et al*. 1998; Semrud-Clikeman 2000; Brewer *et al*. 2001; Willcutt *et al*. 2001). The other two studies (Houghton *et al*. 1999; Klorman *et al*. 1999) also compared the ADHD predominantly inattentive type and ADHD combined type. Both studies found more impairment on the WCST in the ADHD combined type. Taken together, these studies show that both the Tower of London/Hanoi tasks and the WCST capture core deficits in ADHD.

ADHD and Inhibition

Inhibition is currently seen as the core deficit in ADHD; as Barkley states, 'the essential impairment in ADHD is a deficit involving response inhibition' (Barkley 1997, p. 65). There are already several new reviews that clearly document this impairment in ADHD (Barkley 1997; Oosterlaan, Logan and Sergeant 1998; Nigg 2001). Oosterlaan, Logan and Sergeant (1998) focused explicitly on the Stop Signal Task as a measure of response inhibition in their review of eight studies with ADHD children, comparing them to other psychiatric groups, such as children with conduct disorder or anxious children. They report strong effect sizes for two measures of response inhibition when comparing ADHD with controls: for inhibitory function (IF)-slope (efficiency of the inhibitory mechanism, controlling for differences in mean reaction time), $d = 0.94$, and for stop signal reaction time (SSRT; an estimate of the latency of

Table 10.4. Overview of studies (1997–2001) on planning and set-shifting tasks and ADHD

Study	Population	Tasks	Results
Willcutt *et al.* (2001)	52 ADHD children 48 ADHD and RD 93 RD children 121 Controls	WCST CNT	ADHD, RD, ADHD + RD < controls on perseverative errors ADHD + RD, RD < controls on CNT errors
Brewer *et al.* (2001)	26 ADHD children 25 Children with SH 22 Controls	WCST	ADHD < SH, controls on perseverative errors
Semrud-Clikeman (2000)	10 ADD/H children 11 Controls	WCST	ADHD < controls on loss of set. No difference on perseverative errors
Kempton *et al.* (1999)	15 Unmed. ADHD children 15 Med. ADHD children 15 Controls	TLT	Unmed. ADHD < med. ADHD, controls on number of minimum move solutions and total moves in excess of the minimum
Houghton *et al.* (1999)	32 ADHD-I children 62 ADHD-C children 28 Controls	WCST TLT	ADHD-C < controls on categories completed and perseverative errors No differences on Tower of London
Klorman *et al.* (1999)	207 ADHD-C children 102 ADHD-I children Non-ADHD children: 50 Children with other disorders (ODD/RD) 28 Controls	WCST THT	No differences between ADHD types and non-ADHD on perseverative errors ADHD-C < ADHD-I on non-perseverative errors ADHD-C < non-ADHD on solutions and rule violations on THT
Korkman *et al.* (1998)	51 ADHD children Control sample from standardisation group	TLT	ADHD < controls
Aman *et al.* (1998)	22 ADHD children 22 Controls	THT	ADHD < controls on total score, especially when off medication
Pineda *et al.* (1998)	62 ADHD children 62 Controls	WCST	ADHD < controls on all WCST variables
Culbertson and Zillmer (1998)	99 ADHD children 56 Controls	TLT	ADHD < controls on solutions, rule and time violations

Abbreviations: (un)med, (un)medicated; ADHD or ADD/H, attention-deficit/hyperactivity disorder; ADHD-I, attention-deficit/hyperactivity disorder predominantly inattentive type; ADHD-C, attention-deficit/hyperactivity disorder combined type; RD, reading disability/disorder; ODD, oppositional defiant disorder; SH, shunted hydrocephalus; WCST, Wisconsin Card Sorting Test; TLT, Tower of London Test; THT, Tower of Hanoi Test; CNT, Contingency Naming Test.

the inhibitory process) $d = 0.64$. No significant differences emerged between pure ADHD and ADHD co-morbid with conduct disorder on these measures.

Nigg (2001), in his comprehensive review, distinguishes between three types of inhibition: executive inhibition ('deliberate suppression of a cognition or response to achieve a later, internally represented goal', p. 576), motivational inhibition ('cessation of response or behaviour driven substantially by anxiety, uncertainty, or fear', p. 576) and automatic inhibition (relatively automatic suppression of motor or cognitive responses like prepulse startle, inhibition of return or negative priming).

Focusing on executive and motivational inhibition, Nigg (2001) concludes that there is greater support from the literature for a deficit in executive inhibition than in motivational inhibition. In the realm of executive inhibition tasks there are clear and replicated deficits in ADHD on behavioural inhibition tasks such as the Stop task. The evidence on other executive inhibition tasks, such as interference control (e.g. Stroop task), cognitive inhibition (e.g. directed forgetting task) and oculomotor inhibition (e.g. anti-saccade task) is still controversial.

Concerning motivational inhibition, Nigg (2001) also reports conflicting findings for physiological and reward response measures, and reports a preponderance of negative findings for punishment responses measures.

In sum, all three reviews come to the same conclusion, that executive inhibition, especially as seen in the Stop Task, is a or *the* core deficit in ADHD.

THE INTERSECTION BETWEEN ToM, EXECUTIVE FUNCTION AND ADHD

So far, our review has shown that processing ToM tasks is predominantly located in the frontal lobes. Brain imaging studies indicate medial BA8 and BA9 (predominantly left side) and anterior cingulate cortex as most and uniquely activated in ToM tasks. The area most centrally involved seems to be the anterior paracingulate cortex, which forms the layer between the anterior cingulate cortex and BA8 and BA9. As the frontal lobes are also the seat of other higher cognitive functions, especially executive control, the question of which other cognitive tasks activate the same brain areas as ToM tasks arose. From other brain-imaging studies we concluded that self-referential mental activity tasks, including emotional processing, are also activated in these regions. With regard to executive functions, the situation is more controversial. Working memory and problem-solving tasks (e.g. the Wisconsin Card Sorting Test and the Tower of London/Hanoi tests) are primarily activated in the dorsolateral prefrontal cortex and, hence, distinct from ToM areas. In contrast, in conflict and competition tasks (Go/No-Go tasks and especially Stroop-like tasks) we see activations in the medial prefrontal cortex predominantly in the anterior cingulate cortex.

Consequently, the interrelations between these executive function tasks and ToM tasks becomes of crucial importance, and several developmental studies have been undertaken to clarify this relationship. These studies have found significant correlations, in particular between inhibition and working memory tasks and ToM performance. The causal relations are, however, still open to debate, especially for the working memory and problem-solving tasks, as they show activations in regions other than ToM tasks. A key for clearing up these questions would be to find clinical patients who have clear deficits in executive functioning but no deficits in ToM, or vice versa. People with autism or schizophrenia are not good candidates for this purpose, as they show deficits in both. In contrast, ADHD patients seem to fit the bill. Brain-imaging studies indicate that a frontostriatal network is responsible for the core deficits in ADHD, and some studies report specific activation abnormalities in those brain areas that are involved in ToM processing. Neuropsychological studies with ADHD children give an even clearer picture of deficits in executive functions, such as working memory, planning and set-shifting abilities and — especially — inhibition. With this strong evidence for executive problems in children with ADHD, the interesting question is whether they have also problems with ToM, akin to patients with autism or schizophrenia. We are aware of only three published studies investigating ToM abilities in ADHD children or children at risk of developing it (Hughes, Dunn and White 1998; Buitelaar *et al.* 1999; Charman, Carroll and Sturge 2001) and two studies of children with conduct disorder (Happé and Frith 1996; Speltz *et al.* 1999), the major reported co-morbidity in children with ADHD (see Pliszka, Carlson and Swanson 1999).

Hughes, Dunn and White (1998) investigated so-called 'hard-to-manage (H2M)' preschoolers who were rated as hyperactive by parents and teachers (80% of them also had clinically elevated ratings for conduct disorder). There was no clear difference between the clinical and normal preschoolers on false-belief prediction and deception tasks. A difference emerged only on prompted (not spontaneous) explanations, but this was attributable to lower verbal ability in the clinical group. In contrast, the deficits in executive functioning were more pronounced, with four of six tasks revealing significant deficits in the clinical group, in particular working memory and planning (Tower of London test), inhibitory control (Luria's hand game, Detour-reaching box) and attention flexibility (Marbles task). The detour-reaching box results remained significant even when vocabulary scores, fathers' occupational status and mothers' educational status were partialled out.

Charman, Carroll and Sturge (2001) also found no differences in higher-order ToM tasks (Strange Stories from Happé 1994) between ADHD children (6–10 years old) and controls. Significant differences emerged on only one of two executive function tasks (Go/No-Go task, but not Tower of London/Hanoi test). The ADHD group committed more commission errors on the Go/No-Go task than the control group. However, Buitelaar *et al.* (1999)

reported for a subsample in their study significant differences on second-order ToM tasks between ADHD and control children. However, the ADHD subsample was small ($n = 9$) and had unusually severe attention problems, compared even to an autistic group. No specific executive function tasks were employed in this study.

One of the two studies on children with conduct disorders (Happé and Frith 1996) found no deficit at all in children with conduct disorder on first-order false-belief tasks (which seems not too surprising, because the mean age was 9 years). Speltz *et al*. (1999) did not use the usual false-belief tasks as a ToM measure but investigated children's ability to identify positive and negative emotions in cartoon drawings. In their mixed sample (oppositional defiant disorder, with and without ADHD and other disorders) of preschool boys they found significant impairment in the clinical group compared with controls. On two executive function measures the clinical group had a deficit only on the semantic fluency task (but not on motor planning).

We have recently undertaken a study on preschool children (Perner, Kain and Barchfeld 2002), in which nursery teachers assessed 234 children aged 4.5–6.5 years on the DSM-IV criteria for hyperactivity–impulsivity and ADHD. Using slightly lenient classification criteria, 21 children scored on hyperactivity–impulsivity and/or attention deficit and were classified as 'at risk of ADHD'. An age-matched control group showed significant relations between acquisition of an advanced ToM, in particular understanding second-order beliefs, and executive competence as measured by several tasks from the NEPSY. The group at risk of ADHD showed impaired performance relative to the control group on several executive tasks, in particular significantly more misses on a visual attention task, greater problems on the Tower of London test, less fluency and greater distractibility (remain like a statue), but no impairment at all on the advanced ToM tasks. This confirms the impression from the other studies that children with, or at risk of, ADHD have clear executive deficits but no clear deficits in ToM. In any case, the executive deficits seem more pronounced than any ToM deficit that might be there.

CONCLUSION

Our review of brain imaging and neuropsychological studies on ToM and children with ADHD leads to important suggestions and raises interesting questions. The neurophysiological evidence for ADHD is quite clear on an impairment in the fronto-striatal network. It is also clear that the deficit is fronto + striatal, and not just striatal. However, the evidence on where in the prefrontal cortex the deficits and abnormalities are located is far from clear (see also Baumeister and Hawkins 2001). One possibility is that it is a general frontal malfunctioning. This possibility is, however, not tenable in view of the

clear evidence that medial prefrontal cortex (anterior cingulate/paracingulate cortex; see Table 10.1) is central to ToM and that ADHD children are hardly impaired in their ToM (Hughes, Dunn and White 1998; Charman, Carroll and Sturge 2001; Perner, Kain and Barchfeld 2002; Kain, Perner and Traitinger, in preparation). It would indeed imply the absurdity alluded to in our title that children with ADHD do not need their frontal cortex for ToM.

Lacking clear physiological evidence of where to locate the frontal deficit in ADHD, we turn to neuropsychological evidence, which suggests clear deficits in four types of executive tasks whose brain circuits have been investigated. We start with those functions, whose known brain sites are distant from the medial areas involved in ToM processing. Children with ADHD have well-documented deficits in planning (mostly Tower of London Test; see Table 10.4; Pennington & Ozonoff 1996) which involves dorsolateral prefrontal cortex — BA9, BA46 (Dagher *et al.* 1999). They have problems in working memory (see Table 10.3), which is also located in dorsolateral prefrontal cortex (Collette and Van der Linden 2002). Moreover, set-shifting is a problem (see Table 10.4), which also tends to be located dorsolaterally (Frith, Gallagher and Maguire, in press).

On this evidence it appears that the frontal impairment of children with ADHD is located dorsolaterally, leaving their medial prefrontal cortex intact. This could explain why ADHD has a marked impairment on executive functions but not on ToM. Unfortunately for this theory, ADHD also comes with an impairment on inhibition tasks (Go/No-Go and Stroop; Barkley 1997; Oosterlaan, Logan and Sergeant 1998; Nigg 2001), which tend to activate medial areas (anterior cingulate; Barch *et al.* 2001). Although this medial area (rostral posterior cingulate; cognitive division of anterior cingulate, Bush *et al.* 1999) is not the same as activated in ToM tasks (rostral anterior cingulate; Gallagher and Frith in press), it is anatomically very close. This leaves us with the surprising conclusion that ADHD comes with a deficit in one part of the anterior cingulate cortex, next to a seemingly unimpaired region.

ACKNOWLEDGEMENTS

We acknowledge generous financial assistance for the first author from the Austrian Science Fund (FWF Project P13522-SOZ) and invaluable advice from Chris Frith on the most recent brain imaging studies.

REFERENCES

Aman, C.J., Roberts, R.J. Jr and Pennington, B.F. (1998). A neuropsychological examination of the underlying deficit in attention deficit hyperactivity disorder: Frontal lobe vs. right parietal lobe theories. *Dev. Psychol.* **34**, 956–969.

Audenaert, K., Lahorte, P., Brans, B., van Laere, K. *et al.* (2001). The classical stroop interference task as a prefrontal activation probe: a validation study using $^{99}Tc_m$-ECD brain SPECT. *Nuclear Med. Commun.* **22**, 135–143.

Bach, L., Davies, S., Colvin, C., Wijeratne, C. *et al.* (1998). A neuropsychological investigation of theory of mind in an elderly lady with frontal leucotomy. *Cogn. Neuropsychiat.* **3**, 139–159.

Bach, L.J., Happé, F., Fleminger, S. and Powell, J. (2000). Theory of mind: independence of executive function and the role of the frontal cortex in acquired brain injury. *Cogn. Neuropsychiat.* **5**, 175–192.

Barch, D.M., Braver, T.S., Akbudak, E., Conturo, T. *et al.* (2001). Anterior cingulated cortex and response conflict: effects of response modality and processing domain. *Cerebr. Cortex* **11**, 837–848.

Barkley, R.A. (1997). *ADHD and the Nature of Self-control.* Guilford, New York.

Barkley, R.A., Grodzinsky, G. and DuPaul, G.J. (1992). Frontal lobe functions in attention deficit disorder with and without hyperactivity: a review and research report. *J. Abnorm. Child Psychol.* **20**, 163–188.

Barnett, R., Maruff, P., Vance, A., Luk, E.S. *et al.* (2001). Abnormal executive function in attention deficit hyperactivity disorder: the effect of stimulant medication and age on spatial working memory. *Psychol. Med.* **31**, 1107–1115.

Baron-Cohen, S., Tager-Flusberg, H. and Cohen, D.J. (eds) (2000). *Understanding Other Minds. Perspectives from Developmental Cognitive Neuroscience.* Oxford University Press, Oxford.

Baron-Cohen, S., Ring, H., Moriarty, J., Schmitz, B. *et al.* (1994). Recognition of mental state terms. *Br. J. Psychiat.* **165**, 640–649.

Baron-Cohen, S., Ring, H.A., Wheelwright, S., Bullmore, E.T. *et al.* (1999). Social intelligence in the normal and autistic brain: an fMRI study. *Eur. J. Neurosci.* **11**, 1891–1898.

Baron-Cohen, S., Jolliffe, T., Mortimore, C. and Robertson, M. (1997). Another advanced test of theory of mind: evidence from very high functioning adults with autism or Asperger's syndrome. *J. Child Psychol. Psychiat.* **38**, 813–822.

Baumeister, A.A. and Hawkins, M.F. (2001). Incoherence of neuroimaging studies of attention deficit/hyperactivity disorder. *Clin. Neuropharmacol.* **24**, 2–10.

Berthoz, S., Artiges, E., Van de Moortele, P.-F., Poline *et al.* (2002). Effect of impaired recognition and expression of emotions on frontocingulate cortices: an fMRI study of men with alexithymia. *Am. J. Psychiat.* **159**, 961–967.

Blackwood, N.J., Howard, R.J., Bentall, R.P. and Murray, R.M. (2001). Cognitive neuropsychiatric models of persecutory delusions. *Am. J. Psychiat.* **158**, 527–539.

Blair, R.J.R. and Cipolotti, L. (2000). Impaired social response reversal: a case of 'acquired sociopathy'. *Brain* **123**, 1122–1141.

Brewer, V.R., Fletcher, J.M., Hiscock, M. and Davidson, K.C. (2001). Attention processes in children with shunted hydrocephalus versus attention deficit-hyperactivity disorder. *Neuropsychology* **15**, 185–198.

Brunet, E., Sarfati, Y., Hardy-Baylé, M.-C. and Decety, J. (2000). A PET investigation of the attribution of intentions with a non-verbal task. *NeuroImage* **11**, 157–166.

Buitelaar, J.K., van der Wees, M., Swaab-Barneveld, H. and van der Gaag, R. (1999). Theory of mind and emotion-recognition functioning in autistic spectrum disorders and in psychiatric control and normal children. *Dev. Psychopathol.* **11**, 39–58.

Bush, G., Luu, U. and Posner, M.I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn. Sci.* **4**, 215–222.

Bush, G., Frazier, J.A., Rauch, S.L., Seidman, L.J. *et al*. (1999). Anterior cingulate cortex dysfunction in attention-deficit/hyperactivity disorder revealed by fMRI and the counting stroop. *Biol. Psychiat*. **45**, 1542–1552.

Cabeza, R. and Nyberg, L. (2000). Imaging cognition II: an empirical review of 275 PET and fMRI studies. *J. Cogn. Neurosci*. **12**, 1–47.

Carlson, S.M. and Moses, L.J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Dev*. **72**, 1032–1053.

Carlson, S.M., Moses, L.J. and Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant Child Dev*. **11**, 73–92.

Casey, B.J., Castellanos, F.X., Giedd, J.N., Marsh, W.L. *et al*. (1997). Implication of right frontostriatal circuitry in response inhibition and attention-deficit/hyperactivity disorder. *J. Am. Acad. Child Adolesc. Psychiat*. **36**, 374–383

Castellanos, F.X. (2001). Neuroimaging studies. In M.V. Solanto, A.F.T. Arnsten and F.X. Castellanos (eds), *Stimulant Drugs and ADHD. Basic and Clinical Neuroscience*. Oxford University Press, Oxford, pp. 243–258.

Castellanos, F.X., Giedd, J.N., Berquin, P.C., Walter, J.M. *et al*. (2001). Quantitative brain magnetic resonance imaging in girls with attention-deficit/hyperactivity disorder. *Arch. Gen. Psychiat*. **58**, 289–295.

Castelli, F., Happé, F., Frith, U. and Frith, C. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage* **12**, 314–325.

Castelli, F., Frith, C., Happé, F. and Frith, U. (2002). Autism, Asperger's syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*, **125**, 1839–1849.

Channon, S. and Crawford, S. (2000). The effects of anterior lesions on performance on a story comprehension test: left anterior impairment on a theory of mind-type task. *Neuropsychologia* **38**, 1006–1017.

Charman, T., Carroll, F. and Sturge, C. (2001) Theory of mind, executive function and social competence in boys with ADHD. *Emot. Behav. Difficulties* **6**, 31–49.

Cohen, N.J., Vallance, D.D., Barwick, M., Im, N. *et al*. (2000). The interface between ADHD and language impairment: an examination of language, achievement, and cognitive processing. *J. Child Psychol. Psychiat*. **41**, 353–362.

Collette, F. and Van der Linden, M. (2002). Brain imaging of the central executive component of working memory. *Neurosci. Biobehav. Rev*. **26**, 105–125.

Corcoran, R. (2000). Theory of mind in other clinical conditions: is a selective 'theory of mind' deficit exclusive to autism? In S. Baron-Cohen, H. Tager-Flusberg and D.J. Cohen (eds), *Understanding Other Minds. Perspectives from Developmental Cognitive Neuroscience*. Oxford University Press, Oxford, pp. 391–421.

Corcoran, R. (2001). Theory of mind and schizophrenia. In P.W. Corrigan and D.L. Penn (eds), *Social Cognition and Schizophrenia*. American Psychological Association, Washington, DC, pp. 149–174.

Culbertson, W.C. and Zillmer, E.A. (1998). The Tower of London task: a standardized approach to assessing executive functioning in children. *Arch. Clin. Neuropsychol*. **13**, 285–301.

Dagher, A., Owen, A.M., Boecker, H. and Brooks, D.J. (1999). Mapping the network for planning: a correlational PET activation study with the Tower of London task. *Brain* **122**, 1973–1987.

de Zubicaray, G.I., Zelaya, F.O., Andrew, C., Williams, S.C. and Bullmore, E.T. (2000). Cerebral regions associated with verbal response initiation, suppression and strategy use. *Neuropsychologia* **38**, 1292–1304.

Davis, H.L. and Pratt, C. (1996). The development of children's theory of mind: the working memory of explanation. *Aust. J. Psychol.* **47**, 25–31.

Ernst, M., Liebenauer, L.L., King, A.C., Fitzgerald, G.A. *et al.* (1994). Reduced brain metabolism in hyperactive girls. *J. Am. Acad. Child Adolesc. Psychiat.* **33**, 858–868.

Ernst, M., Cohen, R.M., Liebenauer, L.L., Jons, P.H. and Zametkin, A.J. (1997). Cerebral glucose metabolism in adolescent girls with attention-deficit/hyperactivity disorder. *J. Am. Acad. Child Adolesc. Psychiat.* **36**, 1399–1406.

Ernst, M., Zametkin, A.J., Matochik, J.A., Jons, P.H. and Cohen, R.M. (1998). DOPA: decarboxylase activity in attention deficit hyperactivity disorder adults. A [fluorine-18] fluorodopa positron emission tomographic study. *J. Neurosci.* **8**, 5901–5907.

Fine, C., Lumsden, J. and Blair, R.J.R. (2001). Dissociation between 'theory of mind' and executive functions in a patient with early left amygdala damage. *Brain* **124**, 287–298.

Fletcher, P.C., Happé, F., Frith, U., Baker, S.C. *et al.* (1995). Other minds in the brain: a functional imaging study of 'theory of mind' in story comprehension. *Cognition* **57**, 109–128.

Frith, C.D., Gallagher, H.L. and Maguire, E.A. (in press). Mechanisms of control. In R.S.J. Frackowiak, R.J. Dolan, K.J. Friston, C.D. Frith and S. Zeki (eds), *Human Brain Function*, 2nd edn. Elsevier, Amsterdam.

Frith, U. and Frith, C. (2001). The biological basis of social interaction. *Curr. Dir. Psychol. Sci.* **10**, 151–155.

Frye, D., Zelazo, P.D. and Palfai, T. (1995) Theory of mind and rule-based reasoning. *Cogn. Dev.* **10**, 483–527.

Gallagher, H.L. and Frith, C.D. (in press). Functional imaging of 'theory of mind'. *Trends Cogn. Sci.*

Gallagher, H.L., Happé, F., Brunswick, N., Fletcher, P.C. *et al.* (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and non-verbal tasks. *Neuropsychologia* **38**, 11–21.

Gallagher, H.L., Jack, A.I., Roepstorff, A. and Frith, C.D. (2002). Imaging the intentional stance. *Neuroimage* **16**, 814–821.

Giedd, J.N., Blumenthal, J., Molloy, E. and Castellanos, F.X. (2001). Brain imaging of attention deficit/hyperactivity disorder. *Ann. N.Y. Acad. Sci.* **931**, 33–49.

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

Goldstein, S. and Goldstein, M. (1998). *Managing Attention Deficit Hyperactivity Disorder in Children: a Guide for Practitioners*. Wiley, New York.

Gordon, A.C.L. and Olson, D.R. (1998). The relation between acquisition of a theory of mind and the capacity to hold in mind. *J. Exp. Child Psychol.* **68**, 70–83.

Happé, F.G.E. (1994). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *J. Autism Dev. Disord.* **24**, 129–154.

Happé, F. and Frith, U. (1996). Theory of mind and social impairment in children with conduct disorder. *Br. J. Dev. Psychol.* **14**, 385–398.

Happé, F., Ehlers, S., Fletcher, P., Frith, U. *et al.* (1996). 'Theory of mind' in the brain. Evidence from a PET scan study of Asperger's syndrome. *NeuroReport* **8**, 197–201.

Happé, F., Brownell, H. and Winner, E. (1999). Acquired 'theory of mind' impairments following stroke. *Cognition* **70**, 211–240.

Happé, F., Malhi, G.S. and Checkley, S. (2001). Acquired mind-blindness following frontal lobe surgery? A single case study of impaired 'theory of mind' in a patient treated with stereotactic anterior capsulotomy. *Neuropsychologia* **39**, 83–90.

Heider, F. and Simmel, M. (1944). An experimental study of apparent behavior. *Am. J. Psychol.* **57**, 243–259.

Houghton, S., Douglas, G., West, J., Whiting, K. *et al.* (1999). Differential patterns of executive function in children with attention-deficit hyperactivity disorder according to gender and subtype. *J. Child Neurol.* **14**, 801–805.

Hughes, C. (1998). Executive function in preschoolers: links with theory of mind and verbal ability. *Br. J. Dev. Psychol.* **16**, 233–253.

Hughes, C., Dunn, J. and White, A. (1998). Trick or treat?: uneven understanding of mind and emotion and executive dysfunction in 'hard to manage' preschoolers. *J. Child Psychol. Psychiat.* **39**, 981–994.

Johnson, S.C., Baxter, L.C., Wilder, L.S., Pipe, J.G. *et al.* (2002). Neural correlates of self-reflection. *Brain* **125**, 1808–1814.

Kain, W., Perner, J. and Traintinger, K. (in preparation). Theory of Mind and Working Memory in ADHD Children. Unpublished manuscript, University of Salzburg.

Karatekin, C. and Asarnow, R.F. (1998). Working memory in childhood-onset schizophrenia and attention-deficit/hyperactivity disorder. *Psychiat. Res.* **80**, 165–176.

Keenan, T. (1998). Memory span as a predictor of false belief understanding. *N. Z. J. Psychol.* **27**, 36–43.

Kempton, S., Vance, A., Maruff, P., Luk, E. *et al.* (1999). Executive function and attention deficit hyperactivity disorder: stimulant medication and better executive function performance in children. *Psychol. Med.*, **29**, 527–538.

Klorman, R., Hazel-Fernandez, L.A., Shaywitz, S.E., Fletcher, J.M. *et al.* (1999). Executive functioning deficits in attention-deficit/hyperactivity disorder are independent of oppositional defiant or reading disorder. *J. Am. Acad. Child Adolesc. Psychiat.* **38**, 1148–1155.

Korkman, M., Kirk, U. and Kemp, S. (1998). *NEPSY. A Developmental Neuropsychological Assessment Manual*. Psychological Corporation, Orlando, FL.

Kuntsi, J., Oosterlaan, J. and Stevenson, J. (2001). Psychological mechanisms in hyperactivity: I. Response inhibition deficit, working memory impairment, delay aversion, or something else? *J. Child Psychol. Psychiat.* **42**, 199–210.

Lane, R.D. (2000). Neural correlates of conscious emotional experience. In R.D. Lane and L. Nadel (eds), *Cognitive Neuroscience of Emotion*. Oxford University Press, New York, pp. 345–370.

Langleben, D.D., Austin, G., Krikorian, G., Ridlehuber, H.W. *et al.* (2001). Interhemispheric asymmetry of regional cerebral blood flow in prepubescent boys with attention deficit hyperactivity disorder. *Nuclear Med. Commun.* **22**, 1333–1340.

Lezak, M.D. (1995). *Neuropsychological Assessment*, 3rd edn. Oxford University Press, New York.

Liddle, P.F., Kiehl, K.A. and Smith, A.M. (2001). Event-related fMRI study of response inhibition. *Human Brain Mapping* **12**, 100–109.

Lou, H.C., Henriksen, L. and Bruhn, P. (1984). Focal cerebral hypoperfusion in children with dysphasia and/or attention deficit disorder. *Arch. Neurol.* **41**, 825–829.

Lou, H.C., Henriksen, L., Bruhn, P., Borner, H. and Nielsen, J.B. (1989). Striatal dysfunction in attention deficit and hyperkinetic disorder. *Arch. Neurol.* **46**, 48–52.

Lough, S., Gregory, C. and Hodges, J.R. (2001). Dissociation of social cognition and executive function in frontal variant frontotemporal dementia. *Neurocase* **7**, 123–130.

McCabe, K., Houser, D., Ryan, L., Smith, V. and Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proc. Natl Acad. Sci. USA* **98**, 11,832–11,835.

Mehta, M.A., Owen, A.M., Sahakian, B.J., Mavaddat, N. *et al.* (2000). Methylphenidate enhances working memory by modulating discrete frontal and parietal lobe regions in the human brain. *J. Neurosci.* **20** (RC65), 1–6.

Menon, V., Adleman, N.E., White, C.D., Glover, G.H. and Reiss, A.L. (2001). Error-related brain activation during a Go/NoGo response inhibition task. *Hum. Brain Mapping* **12**, 131–143.

Milch-Reich, S., Campbell, S.B., Pelham, W.E., Connelly, L.M. and Geva, D. (1999). Developmental and individual differences in children's on-line representations of dynamic social events. *Child Dev.* **70**, 413–431.

Moses, L.J. (2001). Executive accounts of theory-of-mind development. *Child Dev.* **72**, 688–690.

Murphy, K.R., Barkley, R.A. and Bush, T. (2001). Executive functioning and olfactory identification in young adults with attention deficit-hyperactivity disorder. *Neuropsychology* **15**, 211–220.

Nigg, J.T. (2001). Is ADHD a disinihibitory disorder? *Psychol. Bull.* **127**, 571–598.

Øie, M., Sundet, K. and Rund, B.R. (1999). Contrasts in memory functions between adolescents with schizophrenia or ADHD. *Neuropsychologia* **37**, 1351–1358.

Oosterlaan, J., Logan, G.D. and Sergeant, J.A. (1998). Response inhibition in AD/HD, CD, co-morbid AD/HD + CD, anxious, and control children: a meta-analysis of studies with the Stop task. *J. Child Psychol. Psychiat.* **39**, 411–425.

Pennington, B.F. and Ozonoff, S. (1996). Executive functions and developmental psychopathology. *J. Child Psychol. Psychiat.* **37**, 51–87.

Perner, J., Kain, W and Barchfeld, P. (2002). Executive control and higher-order theory of mind in children at risk of ADHD. *Infant Child Dev.* **11**, 141–158.

Perner, J. and Lang, B. (1999). Development of theory of mind and executive control. *Trends Cogn. Sci.* **3**, 337–344.

Perner, J., Leekam, S.R. and Wimmer, H. (1987). Three year-olds' difficulty with false belief: the case for a conceptual deficit. *Br. J. Dev. Psychol.* **5**, 125–137.

Phan, K.L., Wager, T., Taylor, S.F. and Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage* **16**, 331–348.

Pineda, D., Ardila, A., Rosselli, M., Cadavid, C., Mancheno, S. and Mejia, S. (1998). Executive dysfunctions in children with attention deficit hyperactivity disorder. *Int. J. Neurosci.* **96**, 177–196.

Pliszka, S.R., Carlson, C.L. and Swanson, J.M. (1999). *ADHD with Co-morbid Disorders*. Guilford, New York.

Roodenrys, S., Koloski, N. and Grainger, J. (2001). Working memory function in attention deficit hyperactivity disordered and reading disabled children. *Br. J. Dev. Psychol.* **19**, 325–337.

Rowe, A.D., Bullock, P.R., Polkey, C.E. and Morris, R.G. (2001). 'Theory of mind' impairments and their relationship to executive functioning following frontal lobe excisions. *Brain* **124**, 600–616.

Rubia, K., Taylor, E., Smith, A.B., Oksannen, H. *et al.* (2001). Neuropsychological analyses of impulsiveness in childhood hyperactivity. *Br. J. Psychiat.* **179**, 138–143.

Rubia, K., Overmeyer, S., Taylor, E., Brammer, M. *et al.* (1999). Hypofrontality in attention deficit hyperactivity disorder during higher-order motor control: a study with functional MRI. *Am. J. Psychiat.* **156**, 891–896.

Russell, J. (1997). How executive disorders can bring about an inadequate 'theory of mind'. In J. Russell (ed.), *Autism as an Executive Disorder*. Oxford University Press, Oxford, pp. 256–304.

Russell, T.A., Rubin, K., Bullmore, E.T., Soni, W. *et al*. (2000). Exploring the social brain in schizophrenia: left prefrontal underactivation during mental state attribution. *Am. J. Psychiat*. **157**, 2040–2042.

Sabbagh, M.A. and Taylor, M. (2000). Neural correlates of theory-of-mind reasoning: an event-related potential study. *Psychol. Sci*. **11**, 46–50.

Schweitzer, J.B., Faber, T.L., Grafton, S.T., Tune, L.E. *et al*. (2000). Alterations in the functional anatomy of working memory in adult attention deficit hyperactivity disorder. *Am. J. Psychiat*., **157**, 278–280.

Semrud-Clikeman, M., Steingard, R.J., Filipek, P., Biederman, J. *et al*. (2000). Using MRI to examine brain–behavior relationships in males with attention deficit disorder with hyperactivity. *J. Am. Acad. Child Adolesc. Psychiat*. **39**, 477–484.

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

Solanto, M.V., Arnsten, A.F.T. and Castellanos, F.X. (2001). The neuroscience of stimulant drug action in ADHD. In M.V. Solanto, A.F.T. Arnsten and F.X. Castellanos (eds), *Stimulant Drugs and ADHD. Basic and Clinical Neuroscience*. Oxford University Press, Oxford, pp. 355–379.

Spalletta, G., Pasini, A., Pau, F., Guido, G. *et al*. (2001). Prefrontal blood flow dysregulation in drug naive ADHD children without structural abnormalities. *J. Neural Transm*. **108**, 1203–1216.

Speltz, M.L., DeKlyen, M., Calderon, R., Greenberg, M.T. and Fisher, P.A. (1999). Neuropsychological characteristics and test behaviors of boys with early onset conduct problems. *J. Abnorm. Psychol*. **108**, 315–325.

Stone, V.E., Baron-Cohen, S. and Knight, R.T. (1998). Frontal lobe contributions to theory of mind. *J. Cogn. Neurosci*. **10**, 640–656.

Stuss, D.T., Gallup, G.G. Jr and Alexander, M.P. (2001). The frontal lobes are necessary for 'theory of mind'. *Brain* **124**, 279–286.

Surian, L. and Siegal, M. (2001). Sources of performance on theory of mind tasks in right hemisphere-damaged patients. *Brain Lang*. **78**, 224–232.

Vaidya, C.J., Austin, G., Kirkorian, G., Ridlehuber, H.W. *et al*. (1998). Selective effects of methylphenidate in attention deficit hyperactivity disorder: a functional magnetic resonance study. *Proc. Natl Acad. Sci. USA* **95**, 14494–14499.

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S. *et al*. (2001). Mind reading: neural mechanisms of theory of mind and self-perspective. *NeuroImage* **14**, 170–181.

Wellman, H.M., Cross, D. and Watson, J. (2001). Meta-analysis of theory of mind development: the truth about false belief. *Child Dev*. **72**, 655–684.

White, S. and Keenan, T. (unpublished). Theory of Mind and Executive Functioning: a Longitudinal Study.

Willcutt, E.G., Pennington, B.F., Boada, R., Ogline, J.S. *et al*. (2001). A comparison of the cognitive deficits in reading disability and attention-deficit/hyperactivity disorder. *J. Abnorm. Child Psychol*. **110**, 157–172.

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

Zametkin, A.J., Nordahl, T.E., Gross, M., King, A.C. *et al*. (1990). Cerebral glucose metabolism in adults with hyperactivity of childhood onset. *N. Engl. J. Med*. **323**, 1361–1366.

Zametkin, A.J., Liebenauer, L.L., Fitzgerald, G.A., King, A.C. *et al*. (1993). Brain metabolism in teenagers with attention-deficit hyperactivity disorder. *Arch. Gen. Psychiat*. **50**, 333–340.

11

Social Cognition following Prefrontal Cortical Lesions

ROBIN G. MORRIS, JESSICA BRAMHAM AND ANDREA ROWE
Institute of Psychiatry, University of London, UK

Until recently, little was known about the neuropsychological basis for social cognition and how the brain facilitates social interaction, including the acquisition of social knowledge, the perception and processing of social signals and representation of mental states. All this is changing, with an integration of different approaches, genetic, developmental and neuropsychological (Adolphs 2001).

A key early finding is that impairments in social cognition can apparently occur in the absence of deficits in other main aspects of cognitive function. To some extent, this was epitomised by the iconic patient Phineas Gage, who showed largely intact intellectual, language and memory powers but a profound change in social behaviour (Harlow 1848, 1868). With the benefit of more modern and systematic methods for evaluation, this has been shown convincingly in such cases as EVR (Eslinger and Damasio 1985) who, following the neurosurgical removal of a large orbitofrontal meningioma, was rendered incapable of engaging in normal patterns of social behaviour but had preserved social knowledge (Saver and Damasio 1991). Other accounts of single case studies explore the nuances of these social deficits and often highlight subtle indicators of social impairment, including diminished sensitivity to socially relevant stimuli and emotional responsiveness (Nies 1999; Bach *et al*. 2000; Happé, Malhi and Checkley 2001).

These single cases heavily implicate the prefrontal cortex in social cognition and have been supplemented by group studies that add weight to this general conclusion (Blumer and Benson 1975; Stuss and Benson 1984; Grafman *et al*.

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

1986, 1996). Furthermore, the developmental literature and an increased understanding of theoretical constructs relating to social behaviour provide a framework for this type of analysis. One distinction is between 'cold cognitions' and 'hot cognitions'. Cold cognitions entail the representation of mental states, such as beliefs and intentions, encapsulated in theory of mind (ToM) studies. In contrast, hot cognitions involve the ability to process and draw inferences from affective states, such as emotions, preferences and social threat or safety signals (Stone 1999). It is likely that these dissociate neuronally and the distinction is already strongly supported by functional neuroimaging studies. For example, medial frontal activation is specifically associated with ToM (Fletcher *et al.* 1995; Gallagher *et al.* 2000), whereas emotion recognition, such as identifying the emotional expression of faces, is linked to activity of the left inferior prefrontal cortex (Streit *et al.* 1999), in addition to producing amygdala activation (Morris *et al.* 1996; Phillips *et al.* 1997).

To what extent are these neuronal distinctions upheld by systematic experimental study of patients with frontal brain lesions? The purpose of this review is to explore this question, focusing on three main areas that reflect the authors' interests. First, it considers investigations of group studies of ToM. These have used different experimental methods, such as stories, cartoons and perspective-taking techniques, borrowing from studies of autism and Asperger's syndrome, where false belief and perspective-taking tasks have been developed extensively (Baron-Cohen, Tager-Flusberg and Cohen 1999). Second, it reviews investigations of recognition of emotion expression, including facial and vocal expressions. Finally, the effect of prefrontal cortex lesions on degree of insight into social and emotional dysfunction is described, showing how deficits may divide into different facets.

THE PREFRONTAL CORTEX

Studies of prefrontal cortex damage have the dual purpose of exploring the neurobiology of social cognition and the clinical utility of providing information about social cognition impairment in such patients. The clinical background also informs the extent to which inferences can be made about the involvement of different brain regions, based on the site of brain lesion, and this aspect is reviewed in relation to each study. The majority of the studies reviewed below make the distinction between three main regions and so these are outlined as follows:

1. *The dorsolateral prefrontal cortex* (DFC). This is the lateral and mainly upper anterior region [Brodmann areas (BAs) 8, 9, 10, 44, 45 and 46], thought to subserve reasoning, abstract thinkng and problem solving (Fuster 1997).

2. *The orbitofrontal cortex* (OFC). This is the ventral region of the frontal lobes, and comprises a medial area (BA11), a lateral area (BA47) and a polar area (ventral region of BA10). It has been associated specifically with social function, emotion-related interpersonal behaviours and adaptation to changing reward contingencies (Eslinger 1999).
3. *The medial frontal cortex*. This medial area covers a wide region (BA32 and BA12 and the mesial aspects of BA8, BA9 and BA10). Damage has been associated with impaired affect, emotional processing and motivation (Bowen 1989).

THEORY OF MIND AND FRONTAL LOBE BRAIN LESIONS

There is a debate as to whether ToM is supported by a dedicated cognitive mechanism or is part of a general purpose executive system (Leslie and Roth 1993). This would imply that ToM could in principle be impaired specifically by brain lesions, provided their neural substrate was sufficiently distinct from other systems. In order to explore this issue, several studies have been conducted into the effects of damage to different regions of the brain, including the frontal lobes. Here, a distinction can be made between studies of patients who have widespread lesions involving the frontal lobes, here termed *anterior lesions*, and those in which the lesions are circumscribed, termed *focal lesions*.

STUDIES INVESTIGATING PATIENTS WITH ANTERIOR LESIONS

The first published group study to explore the effects of frontal lobe lesions on ToM was conducted by Stone, Baron-Cohen and Knight (1998). They included five patients with bilateral damage to the orbitofrontal cortex (OFC) due to head trauma. Some of these patients also had damage to the temporal lobe, as would be expected with closed head injury. There were also five patients who had left lateral frontal cortex damage that included dorsal regions of the lateral frontal cortex and more ventrolateral regions. These patients had middle cerebral artery strokes. The area of overlap included the middle frontal gyrus and the depth of the middle frontal sulcus. In summary, this group had primarily dorsolateral frontal cortical (DFC) damage, defining their group membership.

The study used tasks which were graded in developmental difficulty: first-order false-belief tasks, known to be sensitive to ToM development at around 4 years; second-order false-belief tasks, which tests ToM ability at a developmental age of 6–7 years. Both of these were based around stories that were read to the participant, supplemented by presentation on video. For both tasks, there was a condition when the questions came after the story, and

a 'no memory load' condition in which video stills of the stories were left in front of the participant during questioning.

There was also a *faux pas* test, which is more demanding of ToM abilities, since it is associated with cognitive development at around 9–11 years. These tests use stories that depict one of the characters unintentionally making a social gaff, e.g:

> Jeanette bought her friend Anne a crystal bowl for a wedding gift. Anne had a big wedding and there were a lot of presents to keep track of. About a year later, Jeanette was over one night at Anne's for dinner. Jeanette dropped a wine bottle by accident on the crystal bowl, and the bowl shattered. 'I'm really sorry, I've broken the bowl,' said Jeanette. 'Don't worry,' said Anne, 'I never liked it anyway. Someone gave it to me for my wedding'.

The participant has to answer the *faux pas* test question, 'Did someone say something they shouldn't have said?', and also the control question, 'What had Jeanette given Anne for her wedding?'.

The OFC patients showed very few errors on the false-belief tasks throughout, suggesting no ToM impairment elicited by these simpler tests. On the *faux pas* tests, however, they were significantly impaired at answering the test question, even though they had no difficulty with control questions. In contrast, the DFC patients were impaired on the false-belief tests, although this impairment was ameliorated in the 'no memory load' condition, suggesting that the impairment was not specific to ToM. On the *faux pas* task they were again impaired, but this was linked also to non-specific errors in processing the stories. Here, they tended to make errors on the *faux pas* questions when they were also confused about the story details.

These results suggest a dissociation between the deficits exhibited by the two patient groups: the OFC group showed ToM impairments, most distinctively on the more demanding *faux pas* tests; the DFC patients showed impairments only when linked to either memory impairment or non-specific story-processing problems. This finding suggests that the OFC group had a specific difficulty in recognising when somebody had said something inappropriate, just as it has been noted that patients with frontal lobe damage may tend to make *faux pas* comments themselves. A further aspect of the study was to test empathic understanding by asking how the main protagonist in the story felt (e.g. Jeanette in the *faux pas* story given above). Notably, the OFC patients were able to answer successfully, even when they had failed the *faux pas* questions. This result may suggest that an affective response is activated, but that it is inadequately integrated with information about mental states (Stone 1999).

Although these findings at first sight support the notion that the ventral prefrontal cortex is specifically involved in ToM, some caution should be exercised, given the differences in aetiology of brain damage between the

groups. The OFC had bilateral lesions, with the likelihood that head trauma would produce brain injury beyond that identified using either CT or structural MRI. The DFC brain damage, due to middle cerebral artery stroke, was unilateral and also the territory of this structure goes beyond the areas of damage identified by either CT or MRI structural neuroimaging, as indicated in the study. For example, two of the patients were characterised as having anomic aphasia. This comparison raises several issues; one is that unilateral lesions may not be sufficient to produce the *faux pas* deficit; alternatively, the left-sided nature of these lesions may have determined the result.

Indeed, recently Stone (1999) has alluded to the possibility of bilateral DFC lesions being necessary for impairment on the basis of two patients who showed deficits on a perspective-taking task involving map reading (see above; Price *et al*. 1990). An alternative interpretation is that the OFC lesions, as well as being bilateral, were more widespread than identified, the effects of head trauma tending to be diffuse as well as focal. If this is so, then the *faux pas* impairment observed by Stone, Baron-Cohen and Knight (1998) could be related more to diffuse damage to a system supporting ToM, rather than to exclusively prefrontal cortex foci.

A second study of patients with anterior lesions was conducted by Channon and Crawford (2000). They assessed patients with damage of mixed aetiology, including vascular damage, tumours, abscess and sclerosis, who were split into left ($n = 6$) and right ($n = 13$) anterior lesion groups. However, although patients with current dysphasic disturbance were excluded, in many instances the patients had additional damage in the posterior cortex, including the temporal and parietal lobes. A posterior lesion control group was used for comparison. In this study, they employed a test of ToM, the Story Comprehension Test, in which a short story was read, ending with a final sentence that required a non-literal interpretation when considered in the context of the passage. For example, the following story was used:

> Marie dreaded her trips to meet her husband's relatives because they were so boring. Most of the time, they all sat in awkward silence, and this occasion was no different. On the way home, Marie's husband asked her how she had found the visit. Marie said 'Oh, marvellous. I could hardly get a word in edgeways.

The participant is asked the question, 'Why did Marie say that?'.

This study revealed impairment in the left anterior group, but not the other patient groups. An analysis of the types of errors made showed that the left anterior group were tending to make literal responses and were less likely to use mental state terms in their answers. Often these patients showed no evidence of awareness that interpreting the scenario was required, with prompting not eliciting any further speculation in relation to the motivation of the relevant character in the story.

A potential caveat of this study is that control questions or stories were not used, so the impairment could have related to the complexity of the material, rather than ToM difficulties *per se*. However, care was taken to exclude patients with current dysphasia and, in addition, a test of sentence comprehension was administered, with no impairment in the left anterior group. Nevertheless, the more widespread lesions in this group again raise the question as to whether prefrontal cortical damage specifically is causing the ToM impairment.

STUDIES USING CIRCUMSCRIBED FRONTAL LESIONS

To circumvent the problem of the diffusivity of lesions, two further studies have used patients with circumscribed lesions: Rowe *et al*. (in preparation), who tested patients with focal neurosurgical lesions; and Stuss, Gallup and Alexander (2001), who included a mixed sample of neurosurgery and non-neurosurgery patients.

Rowe *et al*. (2001) investigated a comparatively large sample of patients with unilateral neurosurgical lesions (15 right, 16 left). The patients were administered first- and second-order ToM false-belief tasks, developed specifically for the study. Relatively complex adult stories were used, to reduce the possibility that ameliorative strategies could be used by patients with ToM impairments. Examples of the stories used and corresponding questions are given in Figure 11.1. In this study, there were ToM questions, those testing the ability to make an inference about the story, and memory control questions.

The ToM questions took the form of asking the subject why a state of affairs is the case (e.g. Why is Ruth puzzled to find so many soap suds?). For the first-order ToM, this should generate belief attributions of the type 'A thinks X' or 'A does not know Y' (e.g. the answer, 'She doesn't know he's already put soap in . . . she thinks he's just loaded it'). For the first-order stories, there was a significant impairment in the frontal group and the impairment was equivalent for both right and left lesions. Although there was a deficit in the left group on the inference questions, the ToM deficit remained robust when this was included in a co-variate analysis. There were no impairments on the fact and memory questions, and this may suggest that the more focal lesions used in this experiment restricted the degree of general cognitive deficit.

A similar pattern emerged for the second-order stories. Here, the structure for the ToM question follows the format, 'A thinks B doesn't know X' or 'A thinks B thinks Y' (e.g. When asked, 'Why does Richard say this?', the correct response would be 'Because he assumes she hadn't got the message, and so he thinks she doesn't know he's been home and started decorating'). Performance of both right and left lesion patients was impaired. Again, the left lesion group were impaired on the inference question, but this did not eliminate the ToM

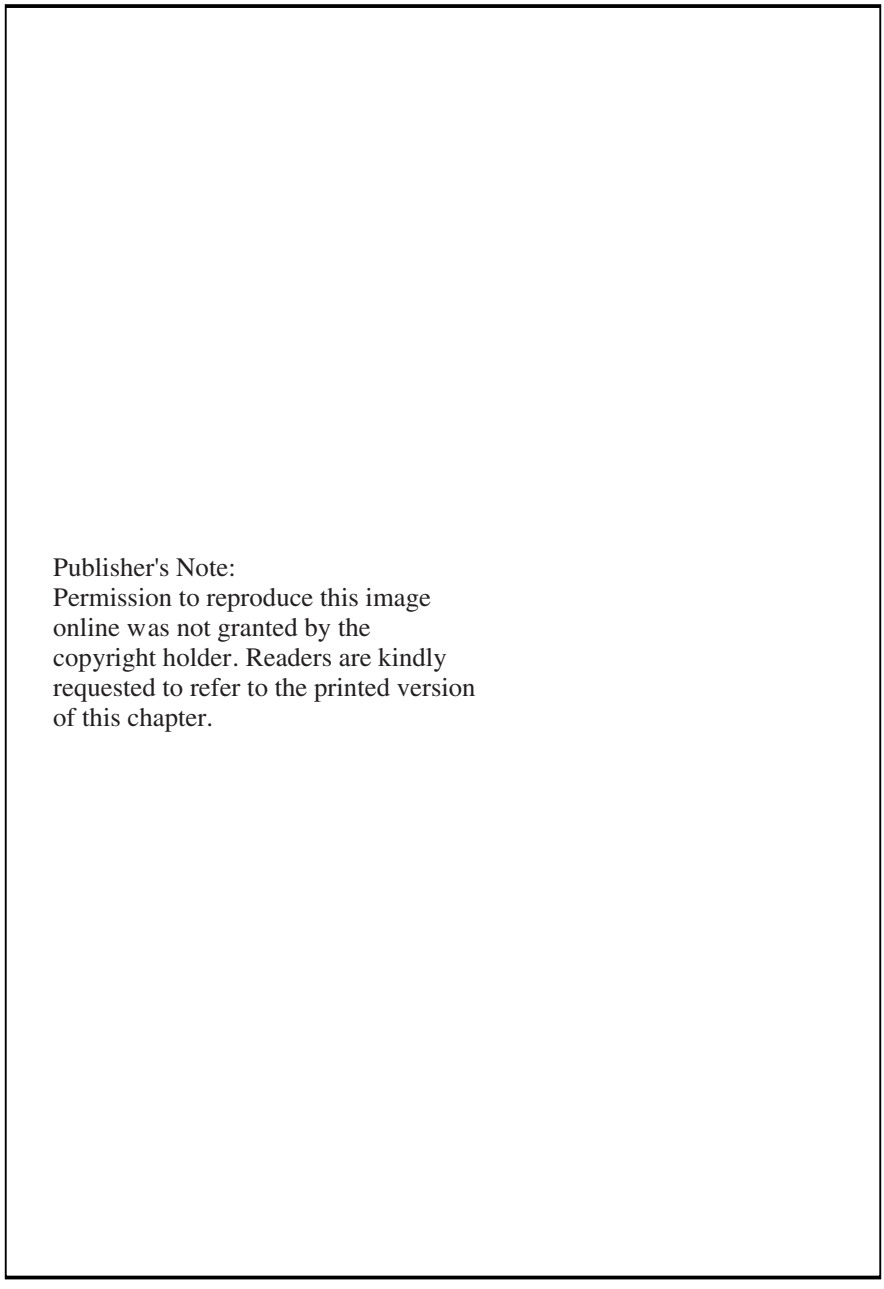


Figure 11.1. Examples of first- and second-order theory of mind stories. Reproduced by permission of Oxford University Press from Rowe *et al.* (2001)

deficit, as indicated by a further co-variate analysis. In addition, the fact and memory questions did not yield deficits in these patients.

A feature of this study was to explore the nature of the participants' responses when they were incorrect. For the first-order stories, patients' errors involved simple inappropriate responses (e.g. 'Because they have put powder in the machine'). These answers tended not to include either incorrect mental state or belief attributions, suggesting that ToM was not being invoked, even in a mistaken fashion. Nevertheless, the second-order stories yielded a somewhat different pattern. Here, there was a tendency for patients with both right and left lesions to make correct first-order belief attributions or inappropriate responses. This finding provides some indication as to the mechanism of ToM failure. It shows, as in the case of the Stone, Baron-Cohen and Knight (1998) study, that the characteristics of a ToM impairment may depend on the procedure used. At first sight it might indicate that in the more complex second-order tasks, the patients were effectively being 'cued' into using ToM, but in a degraded first-order fashion. However, this notion does not stand up to a more detailed analysis of the participants' impairments. Here it was found that those participants who tended to fail on first-order ToM would also fail on second-order ToM stories, and not invoke ToM in their responses. Some patients who passed first-order ToM stories, however, would fail on the second-order ToM stories, and these patients tended to produce the correct first-order belief attributions. Hence, for these patients, there was effectively a degradation of ToM ability elucidated by second-order stories, in which their responses fell back to making only first-order belief attributions.

The results indicated that both left and right prefrontal cortices may contribute to ToM. It was possible to investigate the relative ToM deficit according to whether lesions encroached on either the OFC, DFC or medial frontal cortex. However, this analysis yielded no indication of focal effects, not supporting the distinction made by Stone (1999) between DFC and OFC brain damage.

A second study of ToM involving patients with circumscribed lesions was conducted by Stuss, Gallup and Alexander (2001). The patients had frontal damage of mixed aetiology, including stroke, haemorrhage, lobectomy, tumour and trauma. They were split into groups with unilateral lesions (four right and eight left) and a group with bilateral lesions (seven patients). A group of 13 patients with non-frontal damage and normal controls were used for comparison.

To explore ToM, a visual perspective-taking task was used. This involved the patient and examiner sitting on opposite sides of a table. On the table was a small wooden frame and a curtain which was used as a screen. In addition, there were two assistants who aided in the task presentation and acted as 'stooges' to establish a basis for testing ToM.

The two main conditions were as follows:

1. *Transfer inference*. Here there were five styrofoam coffee cups, situated on the examiner's side of the frame. These were inverted so that they could hide a soft sponge ball underneath one of them. The examiner drew the curtain to hide the cups from the patient and then placed the ball under one of them. One assistant would be on the examiner's side of the table, so that he/she could see where the ball had been hidden. The other assistant was on the participant's side, unable to see. After the ball was hidden, both assistants moved to a set position behind the frame, by the examiner. They then both pointed to a cup, signifying where they thought the ball was located, with the assistant who had been behind the curtain with the examiner pointing to the correct cup. The participant had to infer who was correct, remembering who was positioned originally along side the examiner and judging that this person would know the true location.
2. *Deception condition*. This condition was intended to assess the ability of the participant to infer whether somebody was trying to deceive him/her. Two cups were used to hide a coin and only one assistant was present. This assistant would be with the examiner when the coin was hidden, but subsequently always pointed to the wrong cup. The participant had to infer that the assistant was trying to deceive him/her and so point to the cup that was not selected by the assistant.

There were contrasting results on the two main conditions. An initial analysis of the transfer inference condition for the different groups did not yield a significant difference. However, the frontal patients as a group showed a significant impairment. Also, grouping the right frontal and bilateral patients together revealed an impairment in comparison with non-frontal patients or controls. This was not the case when left frontal and bilateral patients were taken together. In contrast, for the deception condition the bilateral group was impaired, with the left and right frontal groups performing as well as the controls. An analysis of lesion location suggested that involvement of the right medial regions and anterior cingulate was correlated with numbers of errors on this task, the bilateral lesions tending to be medial.

These results point towards a dissociation between perspective taking and judgement of deception, based on brain localisation. However, there are certain issues that need to be addressed before linking the right prefrontal cortex to ToM. First, the number of patients tested with right unilateral lesions was only four, and there remains the possibility that a small sample size incorporates a hidden aetiological bias, the circumscribed nature of the lesion notwithstanding. Second, the visual perspective-taking task involves visuospatial processing, working memory and conditional discrimination. This combination may make patients with right frontal lesions susceptible to impairment for reasons other than problems with perspective taking (Petrides

and Milner 1982; Miotto *et al.* 1996). Stuss, Gallup and Alexander (2001) suggest that the working memory demands are slight and no greater than for another similar condition, which involved the use of opaque or clear glasses to indicate the ability of the assistant to see the location. This conclusion would be strengthened by replicating the finding on a larger group of patients and also by testing spatial working memory and conditional discrimination.

The patients with bifrontal lesions showed a deception impairment, and further lesion analysis indicated that this related to the extent of medial involvement. The lesion analysis does not reveal any distinction between superior and inferior medial damage. However, the result is consistent with the notion that orbital/medial lesions may impair the ability to incorporate perception of deception into the action plan of the patient.

THEORY OF MIND AND THE RIGHT HEMISPHERE

The link between ToM and the right hemisphere has been suggested because of the social and communicative impairments that may accompany right hemisphere lesions, despite there being no aphasic disturbance. This includes problems with processing indirect requests, understanding metaphor and irony and humour, as well as difficulties with cohesive discourse (e.g. Bihrle *et al.* 1986; Brownell *et al.* 1986; Rehak, Kaplan and Gardner 1992; Stemmer, Giroux and Joanette 1994; Van Lancker and Kempler 1997; Winner *et al.* 1998). Although it has been proposed that difficulties in expressing and recognising emotion account for the social impairments (see later), there may also be a link between right hemisphere function and the ability to represent mental states.

Evidence supporting this proposal comes from a study by Siegal, Carrington and Radel (1996), who used very simple false-belief tasks normally used with children. An example is the story, 'Sam wants to find his puppy. Sam's puppy is really in the kitchen. Sam thinks his puppy is in the bathroom', followed by the false-belief question, 'Where will Sam look for his puppy?'. These were administered to left and right hemisphere stroke patients, who had a variety of brain damage in a mixture of frontal, parietal and temporal locations. Only the right group showed impairment, with a high failure rate. Nevertheless, when a control question was administered, e.g. 'Where is it (the puppy) really?', there was also a high failure rate, with a tendency to switch answers and refer to the believed location. Siegal, Carrington and Radel (1996) concluded that this indicated a difficulty with the pragmatic aspects of language. They concluded that the right hemisphere patients may have misinterpreted the control question to mean that they should simply name the location not given in the previous test question. This raises the possibility that failure on ToM may be

secondary to the pragmatic demands of ToM tasks in patients investigated in this study.

Despite this conclusion, the same interpretation does not easily account for the findings of a more extensive study by Happé, Brownell and Winner (1999), again investigating the abilities of patients with right and left hemisphere strokes. They tested 15 right stroke and five left stroke patients using a short passage adapted from those used in functional imaging studies (Fletcher *et al.* 1995; Happé *et al.* 1996). These are of two types, ToM and non-mental stories. Both involve people and having to make inferences, but the first type concerns the characters' thoughts and feelings, whilst the second considers the physical aspects of the story. There was a clear dissociation within the right hemisphere group, whereby they tended to fail on the ToM but passed the non-mental stories. In contrast, the left hemisphere group were not impaired in either condition.

A feature of most ToM tests used with brain-damaged patients is that they are heavily weighted towards language comprehension, whilst in everyday life the substrate for accessing ToM may also involve movement and action of characters. Hence, an alternative approach by Happé, Brownell and Winner (1999) was to use cartoons with humour stemming from either a misunderstanding (ToM) or a physical anomaly (non-mental). For example, one ToM cartoon is a picture of a man and his son in a drawing room. An alien creature on the stairs is not in view of the father but is seen by the son. The caption is the father talking to his son, 'I give up, Robert. What does have two horns, one eye, and creeps?'. An example of a non-mental cartoon is a laboratory scene with one of the characters shrunken, with the caption, 'Looks like Wesselsman hit on something interesting'. They found that the right hemisphere patients showed a robust impairment on the ToM cartoons but performed marginally less well than the control group on the non-mental cartoons. In contrast, the controls tended to perform better on the ToM cartoons, and the left hemisphere group showed no impairment in either condition. A variant on this test was designed with cartoon pairs in which the humorous element of one had been removed and the participant had to select the 'funny' one. This produced a similar pattern of results, with a selective deficit in ToM cartoon choices in the right hemisphere group.

This study points towards a link between ToM and the right hemisphere, consistent across different types of materials. However, a potential caveat is the small sample size of the left group. It is possible, with comparison of left and right stroke patients, that the site and extent of lesion is biased by the need to avoid patients with significant levels of dysphasia, even though there may be matching to lesion site. Also, studying these patients grouped according only to lateralisation raises the question of whether the critical region is across the right hemisphere, involving the frontal, temporal or parietal regions, or is more restricted, for example to the frontal cortex only.

RELATIONSHIP BETWEEN ToM AND EXECUTIVE FUNCTION

A potential explanation for impaired ToM is that it is secondary to executive dysfunction deficits in the control and sequencing aspects of cognition and behaviour. *A priori*, this is plausible because of executive demands made by ToM tasks, such as switching between different types of representations and holding material in working memory. This causal link has been debated in relation to autism and Asperger's syndrome (Ozonoff, Rogers and Pennington, 1991). Here it has been argued further that ToM tests are themselves a type of test of executive function and executive dysfunction in general leads to failure on such tests (Hughes, Russell and Robbins 1994), e.g. one possible mechanism is that the competition between the participant's own knowledge of story details and the inferred false belief of the protagonist may require response inhibition, and failure in this regard leads to an impairment in ToM. Problems with mental flexibility could equally reduce the ability to simultaneously process the story details and the false belief.

In many respects, investigating patients with frontal lesions is a way of testing this link *par excellence* because of known involvement of the prefrontal cortex in executive functioning. The main rationale here is to investigate directly whether there is an association between ToM failure and executive dysfunction. Channon and Crawford (2000) included a battery of executive functioning tests, including Letter Fluency, the Trail Making test, the Hayling Test and the Six Elements test. They grouped all the patients together, including those with anterior and posterior lesions, and found significant correlations between ToM and these tests. They also inspected individual data and found that there were no patients who scored highly on their ToM test and poorly on the executive tests.

This association might point towards a causal link. However, as pointed out above, the lesions in the various groups were not necessarily circumscribed. Since more widespread lesions tend to result in executive dysfunction of greater severity, an association is likely, even in the absence of a causal link. The study by Rowe *et al*. (2001) also employed the Letter Fluency and Trail Making tests, and, in addition, the Stroop test, the Wisconsin Card Sorting test and an externally ordered monitoring task. As would be predicted, deficits were seen on all these tests with frontal lobe lesion patients and there was a tendency for the impairments to occur with left lesions specifically. Rather than correlating these impairments with ToM, there was a reanalysis of the ToM impairments co-varying executive functioning. Thus, if, when the executive functioning tests were co-varied there was still a robust ToM impairment, this would show that the latter was not driven by executive impairment. For both first- and second-order ToM, the ToM impairment remained highly significant. There was some contribution to the variance for some of the executive tests, however, suggesting that executive dysfunction has some influence on performance on ToM tasks.

EMOTION EXPRESSION RECOGNITION AND FRONTAL LOBE BRAIN LESIONS

A further feature of social cognition that can be affected by damage to the prefrontal cortex is the ability to recognise emotions in others. A large component of social communication involves recognition and interpretation of non-verbal information. In particular, the face provides important cues as to others' states of mind and can be regarded as a 'window onto emotion' (Landis 1924). Circumscribed brain damage can impair recognition of facial emotional expression, with the ability to recognise other facial features intact, such as identity or gender (Adolphs *et al.* 1994) and vice versa (Tranel, Damasio and Damasio 1988). Various brain regions have been implicated in recognition of emotion, predominantly in the right hemisphere (Cicone, Wapner and Gardner 1980; De-Kosky *et al.* 1980), including the amygdala (Adolphs *et al.* 1994; Young *et al.* 1995), the insula (Calder *et al.* 2000) and the prefrontal cortex.

Evidence indicating a role for the prefrontal cortex in facial emotion expression recognition comes from non-human primate research (Rolls 1992), neuroimaging studies (George *et al.* 1993; Streit *et al.* 1999), studies with patients with frontotemporal dementia (Fernandez-Duque and Black 2002) and both diffuse and discrete brain lesions (Prigatano and Pribram 1982; Hornak, Rolls and Wade 1996; Rowe *et al.* in preparation). The standard stimuli for assessing facial emotion expression recognition are a series of faces devised by Ekman and Friesen (1975). They prepared photographs of actors displaying seven facial emotion expressions: sad, angry, frightened, disgusted, surprised, happy and neutral. The paradigm generally used involves presenting stimuli individually and asking the subject which of the seven labels best describes the facial expression in the photograph. However, this task has been altered and developed in some studies to involve matching faces by expression or choosing an expression according to a verbal prompt (e.g. George *et al.* 1993; Rowe *et al.* in preparation).

Prigatano and Pribram (1982) used Ekman and Friesen (1975) stimuli to examine recognition and recall of emotional expression of patients with frontal and posterior lesions of mixed aetiology (cerebrovascular accident, tumour, head injury). They found that patients with both bilateral and unilateral frontal lesions were impaired in their recall of facial emotion expressions, whereas patients with posterior damage were significantly less able to recognise emotional expression than normal controls. However, there were no significant differences between the frontal and posterior groups in emotion recognition, except that the posterior group was significantly worse than the frontal group at recognising fear. These results indicate that both frontal and posterior brain regions have some involvement in facial emotion expression processing. This proposition is supported by a magnetoencephalography study by Streit *et al.* (1999) that allowed evaluation of the time course of activation of the neural

network in emotion expression recognition. Earliest activation was seen in the posterior right temporal cortex and inferior occipitotemporal cortex. The middle sector of the temporal cortex (right, then left) was then activated followed by the amygdala, the right anterior cingulate and finally the left inferior prefrontal cortex.

Dysfunction of different neural systems or information-processing strategies may lead to the same impairment in facial emotional expression recognition, but for different reasons. For example, the frontal areas may be more involved in decision-making regarding which emotion is represented (e.g. Adolphs *et al.* 2000) or may be involved specifically in evaluating the emotional content of the faces for learning and shaping behaviour (e.g. Rolls 1996). In contrast, posterior regions may mediate the discrimination of visuospatial aspects of the face.

Deficits in facial emotional expression recognition independent of any perceptual impairments have been shown by patients with ventral frontal lobe damage arising from head injuries and cerebrovascular accidents (Hornak, Rolls and Wade 1996). Patients were also assessed for their ability to recognise vocal emotional expressions. The ventral group was impaired on this task but their deficits did not necessarily concur with poor performance on the facial expression recognition task, indicating some dissociation of auditory and visual emotional expression recognition. In addition, when patients were asked about alterations in their subjective experience of emotion, a strong correlation emerged between the degree of change and their performance on the emotion recognition tasks. Also correlated with subjective change in emotional experience was the severity of informant-rated behavioural problems, such as disinhibition and impulsiveness. Hornak, Rolls and Wade (1996) suggest this may indicate a more general disruption of emotional functioning in patients with ventral lesions, which impacts heavily on their social behaviour.

Nevertheless, although this study divided the patients according to lesion site with more specificity than previous studies, some patients had large lesions due to head injuries and may have incurred diffuse axonal damage not restricted to the prefrontal cortex. In Prigatano and Pribram's (1982) study, a comparison of groups with different lesion aetiology revealed that patients with closed head injuries were significantly more impaired in their perception and recall of emotion expressions than cerebrovascular accident and tumour patients, regardless of the lesion site. Indeed, a recent study of a traumatic brain injury group by Turner, Green and Thompson (2002) showed impaired perception of facial emotional expression, which they interpreted to indicate a role for diffuse axonal injury in the deficit. Evidence for more specialised involvement of the prefrontal cortex has been provided by Rowe *et al.* (in preparation) in a comparison of patients with focal unilateral neurosurgical lesions (see also Morris *et al.* 2002).

Rowe *et al.* (in preparation) used three experimental tasks. Patients were asked to match a target emotional expression to one of six basic expressions. In

a second task, patients were asked to point to an expression, given six different emotions to choose from, following a verbal prompt. Finally, they had to match sketches that depicted emotionally charged scenes in which a protagonist had no facial features with an appropriate facial emotion. When considering the total scores for all three tasks, patients with prefrontal cortex damage were impaired in comparison with normal controls. On the perceptual categorisation task, both left and right groups were less able to match sad faces in particular. For the verbal identification task, both left and right lesions lead to significant impairment in identifying sadness, anger and disgust emotional expressions. For the non-verbal semantic processing task, only the left frontal group were impaired on the sad emotions, and both right and left were impaired on the angry emotions. However, when further group comparisons were made according to extent and location of lesion within the prefrontal cortex, there were no significant differences between the groups. It is possible that the lesions may not have been sufficiently circumscribed to detect differences. Alternatively, facial emotional processing may involve a more extensive cortical network within and beyond the prefrontal cortex, and hence the lesions act to disrupt the functional circuitry more generally.

To date, most studies investigating the role of the prefrontal cortex in facial emotion expression recognition have considered all the basic emotions (sadness, anger, happiness, fear, surprise, disgust) together. However, studies that have examined ability to recognise separate emotions indicate that there may be some fractionation of deficits (e.g. Rowe *et al*. in preparation; Blair and Cipolotti 2000), in which case, it is possible that, rather than having a general role in processing emotional expression information, the prefrontal cortex hosts several circuits mediating separate basic emotional functions.

INSIGHT INTO SOCIAL DEFICITS

A final component of social functioning that has been shown to be altered by prefrontal cortex damage is the capacity for insight. Prigatano, Altman and O'Brien (1990) report that traumatic brain injury patients' judgements of competency in activities of daily living are generally in line with their relatives' ratings. However patients were found to underestimate their deficiencies in emotional and interpersonal interactions. Hornak, Rolls and Wade (1996) found that although patients with ventral lesions were able to appreciate to some extent that they had difficulty in controlling their behaviour and its consequences, the patients were predominantly unaware of their difficulties in recognising emotional expression in others. Instead, they felt that others misinterpreted them and were generally unconcerned about, or underestimated, the seriousness of their condition.

Table 11.1. Subscales of Informant Version of Socio-Emotional Questionnaire

Emotional empathy subscale
- When others are sad, he/she comforts them
- When others are frightened, he/she reassures them
- When others are angry, he/she calms them down
- When others are happy, he/she is pleased for them

Emotion recognition subscale
- He/she notices when other people are angry
- He/she notices when other people are happy
- He/she notices when other people are sad
- He/she notices when other people are frightened
- He/she notices when other people are disgusted

Relationship skills subscale
- He/she does what he/she wants to and doesn't care what others think (R)
- He/she speaks his/her mind (R)
- He/she is close to his/her family

Anti-social behaviour subscale
- He/she avoids arguments
- He/she is impatient with other people (R)
- He/she is not aggressive
- He/she is critical of others (R)

Public behaviour
- He/she is confident meeting new people
- He/she expresses his/her feelings appropriately in public
- He/she cooperates with others

Note: Each item is rated on a 1–5 scale (strongly disagree, slightly disagree, in between, slightly agree, strongly agree).
(R) indicates an unsociable item which is scored in reverse for the total.
From Bramham *et al*. (2002).

Bramham *et al*. (2002) investigated the insight of patients with focal neurosurgical prefrontal cortex lesions into social and emotional functioning, using a specially devised questionnaire. This consisted of 20 items concerning social and emotional dysfunction typically associated with frontal lobe damage, such as aggression and tactlessness, in addition to five items assessing empathic capacity and five items assessing emotion recognition. These were rated by the patients themselves and also by someone who knew them well, according to the extent to which they agreed with each characteristic or tendency. A factor analysis of the items for the informant version of the measure revealed five factors, accounting for 60% of the variance. The 19 items loading most highly on each factor were used to form subscales: emotional empathy; emotion recognition; public behaviour; relationship skills; and anti-social behaviour. Table 11.1 shows the items that make up the five subscales.

There were no significant differences between the patient group and control subjects overall on the total score of the questionnaire. However, the informants of patients with lesions including the orbitofrontal cortex rated their difficulties as significantly greater than controls for both public behaviour and antisocial behaviour. In addition, the orbitofrontal group had significantly more informant-rated difficulties with relationship skills than patients without orbitofrontal damage.

In order to evaluate insight into deficits, patient-rated and informant-rated scores were compared. Patients with orbitofrontal damage significantly underestimated their difficulties with relationship skills and tended to have less insight regarding their antisocial behaviour than their informants. In addition, patients with dorsolateral damage, with or without medial prefrontal cortex damage, were found to overestimate their ability to recognise emotions and empathise with others. These results suggest that different regions of the prefrontal cortex may mediate insight into social and emotional functioning in different domains. Given that insight into behaviour is likely to affect social conduct, knowledge of such deficits may be used therapeutically for rehabilitation purposes.

CONCLUSIONS

This review of studies shows fairly convincingly that social cognition, including ToM, emotional expression recognition and insight, can be affected following lesions affecting the prefrontal cortex. The studies also show that these impairments are not necessarily secondary to deficits in other aspects of cognitive function (e.g. Rowe *et al.* 2001) or perceptual deficits (e.g. Hornak, Rolls and Wade 1996).

Lesion studies investigating ToM abilities have shown mixed patterns of results, which may arise from particular biases in patient selection and the assessment methods used. The study by Stone, Baron-Cohen and Knight (1998) indicated that OFC lesions resulted in ToM deficits on a *faux pas* task, but these patients had bilateral lesions as a result of head trauma and it is very likely that a larger region than the OFC was involved. Channon and Crawford (2000) found ToM impairment associated with left anterior lesions, but again, the diffusivity of the lesions could have influenced these results. With more circumscribed lesions, and using a comparatively large sample, Rowe *et al.* (2001) did not find a link between the site of lesion and either first- or second-order ToM impairment. In contrast, Stuss, Gallup and Alexander (2001) found an association between impairments in perspective taking and right prefrontal cortex, but with medial, particularly right ventral frontal, lesions related to impaired detection of deception. Happe, Fownell and Winner (1999) found impairments in ToM associated with more general right hemisphere lesions.

Group studies have also revealed deficits in the ability to recognise facial emotional expressions following both diffuse damage to the ventral frontal lobes (Prigatano and Pribram 1982; Hornak, Rolls and Wade 1996) and focal lesions following neurosurgery (Rowe *et al.* in preparation). Capacity for awareness regarding such difficulties in recognition of emotion seems to be affected following damage to the DFC and medial prefrontal cortex (Bramham *et al.* 2002). Furthermore, the same study showed a lack of insight regarding inappropriate social behaviour, such as antisocial conduct and poor relationship skills in patients with orbitofrontal damage.

The evidence for association of different aspects of social cognition with particular regions of the prefrontal cortex generally mirrors the results of functional neuroimaging. For example, ToM tasks have been associated with activation in the medial prefrontal cortex (Fletcher *et al.* 1995; Gallagher *et al.* 2000) and the orbitofrontal region (Baron Cohen *et al.* 1994), and emotional expression recognition involves the inferior frontal cortex (George *et al.* 1993; Streit *et al.* 1999). Some theorists argue that social abilities such as ToM are specific innate cognitive mechanisms subserved by designated neurological substrates (e.g. Leslie 1987). Nevertheless, taken together, these lesion studies so far do not provide convincing evidence of localisation in either ToM or emotion recognition within specific regions of the prefrontal cortex.

This raises the possibility that social cognition requires such a diverse set of cognitive operations that it draws on multiple brain systems. For example, Frith (1996) has speculated that a suite of frontal cortical modules may support ToM ability, each adapted and specialised to represent different 'propositional' attitudes. Allied to this is the notion that ToM requires various computational abilities, depending on the precise nature of ToM processing, or the tasks used to test ToM. This may well explain the different brain regions implicated in the studies reviewed above. In order to elucidate this issue, in the future, further work is needed to investigate the components of underlying mechanisms of social cognition and to determine the extent which these dissociate in patients with focal frontal lesions.

REFERENCES

Adolphs, R. (2001). The neurobiology of social cognition. *Curr. Opin. Neurobiol.* **11**, 231–239.

Adolphs, R., Tranel, D., Damasio, H. and Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature* **372**, 669–672.

Adolphs, R., Damasio, H., Tranel, D., Cooper, G. and Damasio, A. (2000). A role for somatosensory cortices in the visual recognition of emotion as revealed by three-dimensional lesion mapping. *J. Neurosci.* **20**, 2683–2690.

Bach, L.J., Happé, F., Fleminger, S. and Powell, J. (2000). Theory of mind: independence of executive function and the role of the frontal cortex in acquired brain injury. *Cogn. Neuropsychiat*. **5** (3), 175–192.

Baron-Cohen, S., Tager-Flusberg, H. and Cohen, D.J. (1999). *Understanding Others' Minds: Perspectives from Developmental Cognitive Neuroscience*. Oxford University Press, Oxford.

Baron Cohen, S., Ring, H., Moriarty, J., Schmitz, B. *et al*. (1994). Recognition of mental state terms: clinical findings in children with autism and a functional neuroimaging study of normal adults. *Br. J. Psychiat*. **165**, 640–649.

Bihrle, A. M., Brownell, H. H., Powelson, J. A. and Gardner, H. (1986). Comprehension of humerous and non-humerous materials by left and right brain-damaged patients. *Brain Cognit*. **5**, 399–411.

Blair, R.J.R. and Cipolotti, L. (2000). Impaired social response reversal. A case of neural responses to facial expressions of sadness and anger. *Brain* **122**, 883–893.

Blumer, D. and Benson, D.F. (1975). Personality changes with frontal and temporal lobe lesions. In D.F. Benson and D. Blumer (eds), *Psychiatric Aspects of Neurological Disease*. Grune and Stratton, New York, pp. 151–169.

Bramham, J., Morris, R.G., Hornak, J. and Rolls, E.T. (2002). Emotional and social consequences of orbitofrontal and non-orbitofrontal lesions of the prefrontal cortex. Paper presented at 'Emotions and the Brain', 12th Rotman Institute Conference, Toronto, Canada.

Bowen, M. (1989). Frontal lobe function. *Brain Inj*. **3**, 109–128.

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

Calder, A.J., Keane, J., Manes, F., Antoun, N. and Young, A.W. (2000). Impaired recognition and experience of disgust following brain injury. *Nature Neurosci*. **3** (11), 1077–1078.

Channon, S. and Crawford, S. (2000). The effects of anterior lesions on performance of a story comprehension test: left anterior impairment on a theory of mind-type task. *Neuropsychologia* **38**, 1006–1017.

Cicone, M., Wapner, W. and Gardner, H. (1980) Sensitivity to emotional expressions and situations in organic patients. *Cortex* **16**, 145–158.

De-Kosky, S., Heilman, K.M., Bowers, D. and Valenstein, E. (1980). Recognition and discrimination of emotional faces and pictures. *Brain Lang*. **9**, 206–214.

Dimitrov, M., Phipps., Zahn, T.P. and Grafman, J. (1999). A thoroughly modern Gage. *Neurocase* **5**, 345–354.

Ekman, P. and Friesen, W.F. (1975). *Unmasking the Face*. Prentice-Hall, Englewood Cliffs, NJ.

Eslinger, P.J. (1999). Orbital frontal cortex. Historical and contemporary views about its behavioral and physiological significance. An introduction to special topic papers: Part 1. *Neurocase* **5**, 225–229.

Eslinger, P. J. and Damasio, A R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: patient EVR. *Neurology* **35**, 1731–1741.

Fernandez-Duque, D. and Black, S.E. (2002). Impaired recognition of negative facial emotions in patients with fronto-temporal dementia. Paper presented at 'Emotions and the Brain', 12th Rotman Institute Conference, Toronto, Canada.

Fletcher, P.C., Happé, F., Frith, U., Baker, S.C. *et al*. (1995). Other minds in the brain: a functional imaging study of 'theory of mind' in story comprehension. *Cognition* **57**, 109–128.

Frith, C. (1996). Brain mechanisms for 'having a theory of mind'. *J. Psychopharmacol*. **10**, 9–15.

Fuster, J.M. (1997). *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe*. Lippincott-Raven, Philadelphia, PA.

Gallagher, H.L., Happé, F., Brunswick, N., Fletcher, P.C. *et al*. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and non-verbal tasks. *Neuropsychologia* **38**, 11–21.

George, M.S., Ketter, T.A., Gill, D.S., Haxby, J.V. *et al*. (1993). Brain regions in facial emotion recognition: An ^{15}O PET study. *J. Neuropsychiat. Clin. Neurosci.* **5**, 384–394.

Grafman, J., Vance, S.C., Weingartner, H., Salazar, A.M. and Amin, D. (1986). The effects of lateralized frontal lobes on mood regulation. *Brain* **109**, 1127–1148.

Grafman, J., Schwab, K., Warden, D. and Pridgen, A. (1996). Frontal lobe injuries, violence, and aggression: a report of the Vietman head injury study. *Neurology* **46**, 1231–1238.

Happé, F., Ehlers, S., Fletcher, P., Frith, U. *et al*. (1996). Distinguishing lies from jokes: theory of mind deficits and discourse interpretation in right hemisphere brain damaged patients. *Brain Lang*. **62**, 89–106.

Happé, F., Brownell, H. and Winner, E. (1999). Acquired 'theory of mind' impairment following stroke. *Cognition* **70**, 211–240.

Happé, F., Malhi, G.S. and Checkley, S. (2001). Acquired mind-blindness following frontal lobe surgery? A single case study of impaired 'theory of mind' in a patient treated with stereotactic anterior capsulotomy. *Neuropsychologia*, **39** (1), 83–90

Harlow, J.M. (1848). Passage of an iron rod through the head. *Boston Med. Surg. J*. **39**, 389–393.

Harlow, J.M. (1868). *Recovery after Severe Injury to the Head*. Massachusetts Medical Society, Boston.

Hornak, J., Rolls, E.T. and Wade, D. (1996). Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal damage. *Neuropsychologia* **32**, 477–492.

Hughes, C., Russell, J. and Robbins, T.W. (1994). Evidence for central executive dysfunction in autism. *Neuropsychologia* **32**, 477–492.

Landis, C. (1924). Studies of emotional reactions: II. General behavior and facial expression. *J. Comp. Psychol*. **4**, 447–509.

Leslie, A.M. (1987). Pretense and representation: the origins of 'theory of mind'. *Psychol. Rev*. **94**, 412–426.

Leslie, A.M. and Roth, D. (1993). What autism teaches us about metarepresentation. In S. Baron-Cohen, H. Tager-Flusberg and D.J. Cohen (eds), *Understanding Other Minds: Perspectives from Autism*. Oxford University Press, Oxford, pp. 83–111.

Miotto, E.C., Bullock, P., Polkey, C.E. and Morris, R.G. (1996). Spatial working memory and strategy formation in patients with frontal lobe excisions. *Cortex* **32**, 613–630.

Morris, J.S., Frith, C.D., Perrett, D.I., Rowland, D. *et al*. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature* **383**, 812–815.

Morris, R.G., Rowe, A.D., Bullock, P.R. and Polkey, C.E. (2002). Facial emotion expression perception following neurosurgical prefrontal lesions. Paper presented at 'Emotions and the Brain', 12th Rotman Institute Conference, Toronto, Canada.

Nies, K.J. (1999). Cognitive and social-emotional changes associated with mesial orbitofrontal damage: assessment and implications for treatment. *Neurocase* **5**, 313–324.

Ozonoff, S., Rogers, S.J. and Pennington, B.F. (1991). Executive function deficits in high functioning autistic individuals: relationship to theory of mind. *J. Child Psychol. Psychiat*. **32**, 1107–1122.

Petrides, M. and Milner, G. (1982). Deficits in subject-ordered tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia* **20** (3), 249–262.

Phillips, M.L., Young, A.W., Senior, C., Brammer, M. *et al.* (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature* **389**, 495–498.

Price, B.H., Daffner, K.R., Stowe, R.M. and Mesulam, M.M. (1990). The comportmental learning disabilities of early frontal lobe damage. *Brain* **113**, 1383–1393.

Prigatano, G.P. and Pribram, K.H. (1982). Perception and memory of facial affect following brain injury. *Percept. Motor Skills* **54**, 859–869.

Prigatano, G.P., Altman, I.M. and O'Brien, K.P. (1990). Behavioural limitations that traumatic brain-injured patients tend to underestimate. *Clin. Neuropsychol.* **4**, 163–176.

Rehak, A., Kaplan, J.A. and Gardner, H. (1992). Sensitivity to conversational deviance in right hemisphere-damaged patients. *Brain Lang.* **42**, 203–217.

Rolls, E.T. (1992). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Phil. Trans. R. Soc. Lond. Ser. B* **335**, 11–21.

Rolls, E.T. (1996). The orbitofrontal cortex. *Phil. Trans. R. Soc. Lond. Ser. B* **351**, 1433–1444.

Rowe, A.D., Bullock, P.R., Polkey, C.E. and Morris, R.G. (2001). 'Theory of mind' impairments and their relationship to executive functioning following frontal lobe excisions. *Brain* **124**, 101–117.

Rowe, A.D., Bullock, P.R., Polkey, C.E. and Morris, R.G. (in preparation). Facial emotion expression perception following neurosurgical prefrontal lesions.

Saver, J.L. and Damasio, A.R. (1991). Preserved access and processing of social knowledge in a patient with acquired sociopathy due to ventromedial frontal damage. *Neuropsychologia* **29**, 1241–1249.

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

Stemmer, J., Giroux, F. and Joanette, Y. (1994). Production and evaluation of requests by right hemisphere brain damaged individuals. *Brain Lang.* **47,** 1–31.

Stone, V.E. (1999). The role of the frontal lobes and the amygdala in theory of mind. In S. Baron-Cohen, H. Tager-Flusberg and D.J. Cohen (eds), *Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience*. Oxford University Press, Oxford, pp. 253–273.

Stone, V.E., Baron-Cohen, S. and Knight, R.T. (1998). Frontal lobe contributions to theory of mind. *J. Cogn. Neurosci.* **10** (5), 640–656.

Streit, M., Ioannides, A.A., Liu, L., Wölwer, W. *et al.* (1999). Neurophysiological correlates of the recognition of facial expressions of emotion as revealed by magnetoencephalography. *Cogn. Brain Res.* **7** (4), 481–491.

Stuss, D.T. and Benson, D.F. (1984). Neuropsychological studies of the frontal lobes. *Psychol. Bull.* **87**, 340–343.

Stuss, D.T., Gallup, G.G. and Alexander, M.P. (2001). The frontal lobes are necessary for 'theory of mind'. *Brain* **124**, 279–286.

Tranel, D., Damasio, A.R. and Damasio, H. (1988). Intact recognition of facial expression, gender, and age in patients with impaired recognition of face identity. *Neurology* **38**, 690–696.

Turner, G.R., Green, R. and Thompson, W.F. (2002). Deficits in facial emotion perception in recently traumatically brain injured adults. Paper presented at 'Emotions and the Brain', 12th Rotman Institute Conference, Toronto, Canada.

Van Lancker, D.R. and Kempler, D. (1997). Comprehension of familiar phrases by left- but not by right-hemisphere damaged patients. *Brain Lang*. **32**, 265–277.
Winner, E., Brownell, H., Happe, F., Blum, A. and Pincus, D. (1998). Distinguishing lies from jokes: theory of mind deficits and discourse interpretation in right hemisphere brain-damaged patients. *Brain Lang*. **62**, 89–106.
Young, A.W., Aggleton, J.P., Hellawell, D.J., Johnson, M. *et al*. (1995). Face processing impairments after amygdalotomy. *Brain* **118**, 15–24.

12

Social Cognition at the Neural Level: Investigations in Autism, Psychopathy and Schizophrenia

TAMARA RUSSELL[1] AND TONMOY SHARMA[2]

[1]Institute of Psychiatry, University of London, and [2]Clinical Neuroscience Research Centre, Kent, UK

Abnormalities in social cognition are an important aspect of the cognitive profile seen in many psychiatric disorders. It is, after all, abnormal behaviour in the social domain and day-to-day life that is often the first indication that an individual is experiencing mental health difficulties. In many disorders, cognition in the social domain is linked to functional outcome (Green *et al.* 2000) and thus a better understanding of the behavioural and neural correlates of this ability are essential.

The term 'social cognition' describes a broad range of cognitive domains, which is beyond the scope of this chapter and is covered in more detail elsewhere in this book. Thus, in this chapter we briefly discuss what we feel are the two main components (face processing and theory of mind) that have implications for imaging. A brief description of these abilities in subjects with autism, schizophrenia and psychopathy is given, followed by a review of the current neuroimaging literature exploring social cognition in these groups.

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

WHAT IS SOCIAL COGNITION?

DEFINITIONS OF SOCIAL COGNITION

Fiske and Taylor (1991) have described social cognition as the way in which people make sense of other people and themselves. In a similar vein, Ostrom (1984) defines social cognition as the domain of cognition that involves the perception, interpretation and processing of social information. The term 'social cognition' is a broad concept, incorporating all aspects of social functioning from perceiving social stimuli (e.g. emotion and face processing), to attributional style and theory of mind (ToM). Perception of emotion allows us to navigate through life and react to our surroundings. ToM is a cognitive process that epitomises the concept of social intelligence, as distinct from general intelligence. It is defined as an innate ability to attribute mental states to ourselves and others in order to predict and explain behaviour. These two elements of social cognition will be the focus of this chapter, as the majority of work that has been conducted using functional neuroimaging has investigated these areas.

There is a suggestion that social cognition may be a special domain of intelligence, and this becomes important if we are to consider specific neural circuits relating to social processing. Investigators studying non-human primates have observed that what may appear to be general-purpose cognitive operations (such as association, reasoning by analogy, making inferences) appear to operate most strongly and consistently in the social domain. Cheney, Seyfarth and Smuts (1986, p. 1364) have reported that primates tested in the laboratory often face problems that are remarkably similar to situations in the wild, however, their performance in these two contexts 'often differs strikingly'. They suggest that one explanation for this finding may be that 'selection for intelligence has acted particularly strongly in the social domain'.

Brothers (1990, p. 28) has defined social cognition as 'the processing of any information which culminates in the accurate perception of the dispositions and intentions of other individuals'. She follows the thinking of Gardner (1983), who suggests the following 'signs' of what he deems to be an 'intelligence': an evolutionary history and plausibility; a distinctive developmental history; the existence of prodigies (and conversely, those who have a selective absence); potential isolation by brain damage; and identifiable core operations.

Brothers (1990, p. 28) notes that, while there are a number of species that interact in highly specific ways (e.g. ants and bees), it is only the primate species that have developed 'a unique capacity to perceive psychological facts (dispositions and intentions) about other individuals'. Brothers is the main proponent of the thesis that social cognition has evolved in the human primate species to address specific requirements in the social environment. She argues

that 'it is logical to assume that, together with the development of ever more differentiated signalling devices, such as the expressive primate face, a cognitive apparatus for correct perception and response evolved as well' (Brothers 1990, p. 30). Additionally, she points out that the pressures of living in a large social group, and the increasing competency for deceiving and manipulating one another, 'is in itself a source of ever more intense pressure to develop acute perception of subtle configurations of expression and behaviour' (Brothers 1990, p. 32).

There is fairly stereotyped ontogenetic development of behaviours in the realm of social cognition in the human infant, ranging from basic mechanism of voice and face perception evident from birth, to the more complex development of ToM, developed in most infants by the age of 4. This stereotypy reflects, she suggests 'evidence of an innate neural specialization for social behaviour' (Brothers 1990, p. 32) and a distinctive developmental trajectory.

Similarly, she cites autism as a case of specific and selective abnormal development in this realm. Although the most cited abnormality in autism is the lack of ToM (see Baron-Cohen 1994, 1996), there are other aspects of social cognition, such as gaze behaviour and protodeclarative pointing, that are abnormal in this group. Some aspects of face processing, such as understanding the emotional content of a face, are also impaired. This can often be in the context of normal (or near-normal) functioning in other, non-social domains.

Although autism is the most extreme example of a fault in the 'social cognition' system, there are other conditions where the impairment may exist in a more subtle form. Conditions exists where only one aspect of the module may be impaired, e.g. the case of prosopagnosia, a specific impairment in the recognition of familiar faces (Brothers 1990). Evidence from lesion studies, particularly those following damage to the orbitofrontal region, suggests that it is possible to 'knock out' elements of social functioning. This is observable in the classic case of patient EVR, who had intact verbal knowledge of appropriate social behaviour but was disastrous when operating in real-life social situations (Eslinger and Damasio 1985). Brothers (1990, p. 57) suggests that as a result of his resection of the orbitofrontal cortex EVR has 'lost access to the internal cues which the behaviour of others should generate'. Another syndrome, Capgras syndrome, is also suggested to selectively affect one aspect of social cognition, notably, the 'felt link with the mental life of another' (Brothers 1990, p. 37). In this condition, the patient holds the absolute conviction that a loved one, although looking exactly as usual, has in fact been replaced by an impostor. Lastly, paranoid psychosis is an example of a condition whereby intentions and motivations are grossly misinterpreted and misattributed.

The evidence cited above suggests that social cognition fits the criteria, as outlined by Gardner (1983), for a separate 'intelligence', but what of the neural correlates of these abilities — do they also support this suggestion?

It is only recently that a surge of interest in the neural basis of aspects of social cognition has arisen; neuroscience seems to have neglected an area that has been well documented by behaviourists. One group who have done pioneering work is that led by Dave Perrett. This group have tirelessly explored the neural basis for many aspects of social cognition, primarily employing single-cell techniques in monkey brains (Perrett, Mistlin and Chitty 1987; Perrett *et al*. 1989, 1990a,b, 1992, 1994). Particular interest has been generated by the discovery of cells, in the bank of the superior temporal sulcus of the macaque brain, which are dedicated to face and gaze perception. Additionally, some of these cells appear to code for interactions between agents and objects and may represent the neural basis for the perception of intention (Perrett *et al*. 1990b). The presence of such cells adds weight to the suggestion that the primate brain may have evolved in a specific way to cope with social situations. Similarly, non-invasive imaging methods are also increasingly shedding light on the relevant areas of the human brain required for social cognition, e.g. Hoffman and Haxby (2000) have demonstrated activity in homologous areas of the human brain in response to eye movement, using fMRI.

SOCIAL COGNITION IN AUTISM

Abnormalities in face processing have been widely reported in individuals with autism (Pierce *et al*. 2001; Davies *et al*. 1994). Difficulties have been noted in the perception of facial affect (Hobson, Ouston and Lee 1988), eye gaze perception (Baron-Cohen *et al*. 1999) and also in facial identity discrimination (Tantam *et al*. 1989). Individuals with autism have been described as 'face-inexperienced' (Pierce *et al*. 2001, p. 2059) and this deficit may be one of a series of developmental milestones that these individuals fail to meet. Interestingly, it has been demonstrated that children with autism do not show the usual processing advantage for upright vs. inverted faces (Tantam *et al*. 1989) which suggests that they may be processing these stimuli in a configural rather than a holistic fashion.

Another developmental milestone not met by autistic individuals is the development of ToM, which is apparent in healthy children at about the age of 4. Deficits in this domain have been robustly demonstrated in this group (Baron-Cohen 1996; Mitchell 1997), and this lack of ToM, or 'mindblindness' (Baron-Cohen 1996) leads to immense difficulties operating in the social environment.

SOCIAL COGNITION IN SCHIZOPHRENIA

It is undisputed that patients with schizophrenia have significant problems with interpersonal interactions (Mueser and Bellack 1998; Green *et al*. 2000). These can be manifested as either poor premorbid social functioning, poor social

functioning as the result of an exacerbation of symptoms, or as a 'deficit state' during remitted periods (e.g. anhedonia) (Penn and Corrigan 1997). This type of dysfunction is such an intrinsic part of the disorder that impairments in social functioning are considered among the hallmarks of the disease and are included in the DSM-IV diagnoses (APA 1994).

The literature to date exploring facial emotion processing in this group of patients presents a rather confused picture that still requires clarification. It has variously been suggested that individuals with schizophrenia have a generalised deficit on all facial tasks (Novic, Luchins and Perline 1984; Feinberg *et al.* 1986; Gessler *et al.* 1989; Kohler *et al.* 2000); that they have a specific deficit with respect to recognising emotions on faces (as compared to recognising gender or identity; see Mandal, Panday and Prasad 1998 for a review); that impairments are limited to specific emotions (Edwards *et al.* 2001); or that there is no impairment at all (Bellack, Blanchard and Mueser 1996). Similarly, there is still debate over whether the observed deficits are the result of state or trait phenomena (Lewis and Garver 1995; Walker, Marwit and Emory 1996; Wölwer *et al.* 1996) and what the role of symptoms might be (Lewis and Garver 1995; Schneider *et al.* 1995). The most likely explanation for these discrepant results is the heterogeneity of the samples tested. Increasingly, neuropsychological and social cognition studies (for an example of ToM investigations, see below) have revealed differential abilities dependent on symptomatology, although this type of approach has yet to be consistently adopted in either the face-processing or functional imaging literature. Additionally, methodological limitations relating to the type of stimulus use (posed, still photographs of Caucasian individuals in most instances) may explain the differing results.

ToM deficits in schizophrenia have now been reported in a number of studies (Corcoran, Mercer and Frith 1995; Frith and Corcoran 1995; Corcoran, Cahill and Frith 1997; Sarfati *et al.* 1997a,b, 1999; Doody *et al.* 1998, Drury, Robinson and Birchwood 1998; Pickup and Frith 2001; Russell and Morris 2001). The most striking thing about these findings is that they suggest that it is the predominant symptomatology individuals present with that determines whether or not they are able to accurately attribute mental states to others. Specifically, it is argued that those with negative symptoms are particularly impaired, although this may be due to more general problems with inference *per se*. On a cartoon task requiring the inference of mental states (experimental condition) and physical states (control condition) they are impaired on both conditions. This is in contrast to the paranoid group, who are selectively impaired on those cartoons requiring mental state inference but are able to successfully make inferences based on physical states. Subjects in remission have thus far been shown to have intact ToM. The work of Corcoran and Frith has indicated that ToM deficits appear to be dependent on state, rather than trait. This type of deficit (which remits when symptoms abate) is ideal for

neuroimaging investigations, as it is possible to investigate neural abnormalities during an acute phase and determine whether this normalises when the subject is aymptomatic and is able to successfully complete the task. To the authors' knowledge, this type of longitudinal study has yet to be conducted.

The Role of Symptoms

While symptom-based approaches have dominated some aspects of social cognition research (notably the ToM literature), there are relatively few other studies that have explored the effect of symptomatology on social functioning and cognition. Penn and Corrigan (1997) have argued that the current models of aberrant cognition in schizophrenia focusing exclusively on non-social-cognitive processes do not adequately explain the social impairment and symptomatology seen in the illness. Liddle's (1987) three-factor solution as a description of schizophrenic syndromes has been influential in determining many aspects of current research, yet relatively few studies of social cognition have taken into account the heterogeneity of symptoms seen in the disorder. Those studies that have included information about symptom status and/or included different symptom dimensions as part of their investigations are discussed below.

Liddle (1987) was one of the first authors to evaluate his three symptom dimensions (psychomotor poverty, reality distortion and disorganisation) with respect to social functioning. He reports that negative and disorganised factors were most strongly related to domains of social functioning. Notably, negative symptoms were related to physical anergia, recreational difficulties and impaired relationships with peers and friends. The disorganised cluster was related to poor grooming and hygiene, lack of persistence at work, lack of intimacy and social inattentiveness. Positive symptoms were not significantly related to any of the social functioning items. This finding has been replicated (Brekke *et al*. 1994). Negative and disorganised symptoms were again more related to social functioning than positive symptoms. Both symptom dimensions were found to be related to 'social competence' (as assessed using the Community Adjustment Form). Residual positive symptoms have not been associated with patients' social deficits (De Jong *et al*. 1986; Prudo and Monrow Blum 1987).

The overall picture suggests that negative symptoms are more strongly related to social functioning than positive symptoms, although some studies suggest that disorganised symptoms may also influence social behaviour, to a lesser degree than negative symptoms but to a greater degree than positive symptoms. This echoes the literature on the lack of relationship of positive symptoms in schizophrenia to functional outcome, and provides evidence that social cognition abilities mediate functional outcome.

SOCIAL COGNITION IN PSYCHOPATHIC DISORDER

Lack of empathy is a core feature of psychopathy. Similarly, within the related diagnosis of antisocial personality disorder (DSM-IV; APA 1994), criteria such as a persistent disregard for the feelings of others are included. Psychopathic individuals have been shown to be less emotional than non-psychopaths (Day and Wong 1996) and also demonstrate autonomic hyporesponsivity to sad and fearful faces (Anskiewicz 1979; Blair *et al*. 1996, 1997); however, this is in the context of no significant difference from controls on facial affect recognition tasks (Blair and Cipolotti 2000).

Regarding ToM abilities, this group have been shown to perform these tasks at a level that is comparable to that of non-psychopathic controls and superior to high-functioning autistics (Blair *et al*. 1996). This group of individuals is of interest, as although they seem able to perform emotion recognition tasks and ToM tasks, it is certain from their behaviour that there is aberrant social cognition.

FUNCTIONAL BRAIN IMAGING STUDIES

FUNCTIONAL NEUROANATOMY OF EMOTION

In a meta-analysis of 55 studies, Phan *et al*. (2002) summarise the current literature relating to the functional neuroanatomy of emotion as explored in healthy subjects. Included are studies requiring responses to individual emotions (positive, negative, happiness, fear, anger, sadness, disgust), to different induction methods (visual, auditory, recall/imagery), and to emotional tasks, with and without cognitive demand. They report that the medial prefrontal cortex has a general role in emotional processing; fear specifically engaged the amygdala; sadness was associated with activity in the subcallosal cingulate; emotional induction by visual stimuli activated the occipital cortex and the amygdala; induction by emotional recall/imagery recruited the anterior cingulate and insula; and emotional tasks with cognitive demand also involved the anterior cingulate and insula.

NEUROIMAGING STUDIES OF FACE PROCESSING IN AUTISM

Difficulties in reading facial affect are well documented in individuals with autism (Hobson 1986; Hobson, Ouston and Lee 1988; Bormann-Kischkel, Vilsmeier and Baude 1995) and it has been argued that these individuals are 'face-inexperienced' (Pierce *et al*. 2001). As such, these subjects provide an opportunity to examine what might be happening at a neural level to systems that are known, in healthy subjects, to be dedicated to face processing. Pierce and colleagues report one of the few studies examining the neural correlates of

face processing in a group of seven adult autistics, as compared to healthy adults. The subjects viewed pictures of faces, and blood oxygen level-dependent response (relative to scrambled shapes) was determined in four main regions of interest (ROI): the fusiform gyrus (FG), inferior temporal gyrus (ITG), middle temporal gyrus (MTG) and amygdala. Volumetric analysis of these four regions was also conducted to explore any structural abnormalities in this group. The autistic individuals in this study did not show any significant detriment in performance (accuracy or reaction time) relative to the control subjects for either the face-processing or shape-matching tasks. Structural analysis showed significantly smaller amygdala volumes in the autistic group, with an average reduction of ~15%. Fusiform gyrus volume was also on average ~8% smaller, although this difference was not statistically significantly different from controls. There were no significant differences in the ITG and MTG volumes in this group. Functional analyses of these ROIs indicated that volumes of activation were significantly smaller in the region of the FG bilaterally and the left amygdala for the autism group; with no differences in the ITG and MTG. This reduction in activation could be related to the smaller structural volumes seen in these individuals and correlational analyses suggested this was the case. Of note from this study is the lack of activation in all except one autistic subject in a region consistently shown to be active in normal subjects in response to faces; i.e. the FG. Rather, these autistic individuals maximally activated a number of individual-specific regions in response to the facial stimuli, e.g. frontal cortex, primary visual cortex and cerebellum. Misguided brain growth, coupled with insufficient experience of face processing in these individuals, are cited as the most likely explanations for these findings. An interesting question raised by these data is what might be the neural consequences of a face-processing training or rehabilitation in these individuals—can the usefulness of the more 'traditional' face-processing regions be reinstated following forced exposure or training? This has yet to be explored.

In a slightly larger study (14 subjects), Schultz *et al*. (2000) also reported abnormalities in the FG (in the right hemisphere only) in a group of autistic and Asperger's syndrome patients. Again, using a ROI approach, it was demonstrated that, compared to healthy controls (who uniformly activate the FG), these individuals do not. Significant differences were also seen in the ITG during the discrimination of faces. Again, this was in the context of no significant differences in performance as measured by accuracy.

More recently, Critchley *et al*. (2000) have replicated the findings of Schultz and colleagues. Nine adults with autistic disorder and nine healthy controls were scanned using fMRI while they implicitly and explicitly processed emotional facial expressions. In the explicit task, the subjects indicated whether the face they saw was happy, angry or neural. The faces were presented in 30 s blocks, alternating emotional and neutral presentations. For the implicit task,

the subjects were required to indicate the gender of the face (male/female). Comparing the brain activation across groups and between tasks, the cerebellar vermis, left lateral cerebellum, striatum, paralimbic and limbic areas, left insula and amygdalahippocampal junction, and left MTG regions reached significance. The most notable finding from this study was the lack of left MTG activation in response to the explicit task in the autism group, although it should be noted that behavioural data indicated a significant deficit in this labelling ability in this group.

Studies of autistic individuals seem to point to abnormalities in the region of the brain that in healthy subjects is dedicated to processing information about faces, viz. the FG. It is argued that in these autistic brains, a lack of expertise has led to the utilisation of a more distributed system for processing faces. This is in keeping with the psychological literature, which posits that autistic individuals and those with Asperger's syndrome are more likely to process faces in a configural manner, rather than a holistic manner (Langdell 1978; Hobson, Ouston and Lee 1988).

NEUROIMAGING STUDIES OF FACE PROCESSING IN SCHIZOPHRENIA

Literature relating to the functional imaging of face processing in schizophrenia is currently sparse. Taking a symptom-based approach, Phillips *et al.* (1999) compared the neural activity to negative facial expressions of emotion in a small group of paranoid subjects with that of non-paranoid subjects and controls. The non-paranoid subjects were actively psychotic; however, their delusions were of a non-persecutory nature (e.g. grandiose or bizarre delusions). The experimental stimuli consisted of facial expressions of fear, disgust and anger. The use of these three emotions allowed for the exploration of neural responses to threatening negative emotions (fear and anger) and non-threatening negative emotions (disgust). This design enabled the authors to address the specific prediction that paranoid subjects would show particularly aberrant responses to the threatening stimuli. In a block design paradigm, responses to these emotions were compared with the response to a neutral (in this case mildly happy) face. The subjects were required to make a gender decision about the faces they viewed, and thus were blind to the experimental protocol. Examination of the results from all schizophrenia patients together revealed that, while appropriate insular activity was seen to the disgust faces, the amygdala was not activated for fear faces, and angry faces failed to activate any area to a significant extent. This suggests that, while networks for non-threatening negative faces (i.e. disgust faces) are intact, those mediating neural responses to threatening faces (anger and fear) may be disturbed. Comparison of the paranoid and non-paranoid patients indicated that non-paranoid patients appeared to have a more general reduction in neural activity to both fearful and disgusted faces. Surprisingly, amygdala activity was demonstrated

in response to angry faces (normally seen in response to fearful faces). In response to fearful faces, no amygdala activity was seen; however, it appeared in response to facial expressions of disgust. Although these findings present rather a mixed picture for this group, the results of behavioural data collected outside the scanner suggested that these patients were having difficulty in correctly identifying the three emotions and were frequently confusing them.

What conclusions can we draw from this study? It is the first to explore emotion processing considering different symptom subgroups of patients with schizophrenia. It shows that there are distributed cortical networks that are abnormal in schizophrenia. It indicates that abnormalities exist in a region of the brain that is used by normal subjects in the recognition of negative expressions of emotion; the amygdala. Of particular note is the lack of appropriate response to the threatening facial stimuli in the paranoid group. However, the number of subjects in each of the subgroups ($n = 5$) is small. We also do not know if these are only state-specific abnormalities, as there were no differences in activity in this region between paranoid and non-paranoid groups. The ideal way to answer this question would be to carry out a longitudinal study following-up patients over time and use them as their own controls. Furthermore, it might be possible to examine the effects of different types of antipsychotics on performance and brain activity.

A more subtle way to explore abnormal affective processing is by using mood-induction paradigms. In such a paradigm, rather than asking subjects to recognise an emotion, that emotion is induced in the subject so that he/she experiences the emotion. Schneider and colleagues presented 13 male medicated patients with schizophrenia and matched controls with happy and sad faces, and asked the subjects to try to become happy or sad as they viewed each face (Schneider *et al*. 1998). The subjects were also asked to rate how they felt on a five-point intensity scale (how happy or how sad). A control condition utilised the same stimuli but subjects were asked to make a gender decision about each face. A ROI approach was adopted, with the amygdala, hippocampus, thalamus, anterior and posterior cingulate, orbito-frontal cortex (OFC), dorsolateral pre-frontal cortex (DLPFC), temporal superior cortex, temporal medial cortex, temporal inferior cortex, occipital cortex, precuneus, and cerebellum chosen as specific regions for comparison across groups. These regions were grouped into sub-cortical-limbic, frontal-limbic, temporal and control regions.

During sad mood induction in healthy males, activation in the amygdala was demonstrated. This was not evident in the schizophrenic individuals. Differences between the two groups were only demonstrated in this subcortical-limbic region — all other frontal-limbic, temporal and control regions showed non-significant main effects. The difference in amygdala activation was not due to smaller volume (approximate volumes provided by anatomical T1 images between the two groups were comparable), or to

differences in subjective ratings of negative mood. This study provides further support for the hypothesis that dysfunction in the amygdala contributes to abnormalities in emotion processing in schizophrenia.

NEUROIMAGING STUDIES OF FACE PROCESSING — PSYCHOPATHS

Abnormalities in affective processing are a hallmark of psychopathology. Specifically, difficulties in processing negatively valenced emotional stimuli are thought to underlie some of the symptoms of the disorder, such as lack of empathy. While there are no fMRI studies to date reporting on face processing abilities in this group, one study has demonstrated that these individuals fail to show normal behavioural facilitation and event-related potential differentiation between emotional and neutral words (Williamson, Harper and Hare 1991). Kiehl *et al.* (2001) wished to examine the neural systems underlying abnormal affective processing in this group. To this end, an affective memory task was used, which consisted of eight repetitions of three phases; encoding, rehearsal and recognition (plus rest). Words had to be encoded, rehearsed and recognised, and the subjects were blind to the fact that four of the eight repetitions contained words that were negative in affect. Criminal psychopaths, criminal non-psychopaths and healthy control subjects participated in the study. There were no differences between the three groups in terms of activation related to processing the neutral stimuli. Differences between the psychopaths and the criminal non-psychopaths during the affective condition were seen in the rostral and caudal anterior cingulate, posterior cingulate, left inferior frontal gyrus, right amygdala and ventral stratium, with less activation in the psychopath group. Compared to the healthy controls, there was less activation in the psychopath group in the left amygdala and parahippocampal gyrus and bilateral anterior superior temporal gyrus. During encoding and rehearsal, the psychopath group showed less activity in the right amygdala, while left amygdala differences were seen during recognition. In this group of individuals it is suggested that limbic system dysfunction, particularly in the amygdala, might be responsible for the insensitivity to fear and punishment contingencies. Activation in response to neutral words was greater among the psychopaths in a number of frontal brain regions, including bilateral inferior lateral frontal cortex. The authors suggest that criminal psychopaths may employ non-limbic cognitive strategies to process affective material. This type of distributed system may be similar to that seen in individuals with autism, who also may compensate for limbic structure abnormalities by using alternative brain regions to process the same information. It is noteworthy that both studies report this type of pattern in the context of intact recognition (Kiehl *et al.* 2001; Pierce *et al.* 2001).

SUMMARY — NEUROIMAGING STUDIES OF FACE PROCESSING IN PSYCHIATRIC DISORDERS

Common to all three disorders, it would appear, are abnormalities in the functioning of subcortical limbic structures, most notably the amygdala. Although no studies on face processing *per se* have yet been conducted in psychopathy, the affective memory task revealed differences in activation in the region of the amygalda. It is noteworthy that, for both high-functioning autistic individuals and those with psychopathy, a reduction in amygdala activation was observed, despite intact performance on the task in question (although this was not the case in the Critchley *et al*. 2000 study). In the two schizophrenia studies described, task performance was not assessed, although behavioural data from this group suggests an impairment for emotion recognition. This raises an interesting point relating to task performance and brain activation; in all three groups there is a reduction in activation in the amygdala, although for some it is in the context of good performance and in others, in the context of poor performance. This region of the brain is notoriously hard to image due to its proximity to air spaces (such as the sinuses). This means that the images obtained can sometimes have very poor resolution. Great care needs to be taken in order to obtain good images from this region of the brain.

ToM

A NEURAL SUBSTRATE FOR ToM?

It is unlikely that there is a single brain region responsible for ToM. Recent functional neuroimaging studies of ToM have pointed towards involvement of the left-hemispheric lateral prefrontal and temporal brain regions in the performance of different tasks of mental state attribution (Fletcher *et al*. 1995; Goel *et al*. 1995; Castelli *et al*. 2000; Gallagher *et al*. 2000; Vogeley *et al*. 2001). The imaging data seem to be at odds with the findings of lesion studies which largely implicate the right hemisphere (Siegal, Carrington and Radel 1996; Winner *et al*. 1998; Happé, Brownell and Winner 1999; but see Rowe *et al*. 2001). The majority of the imaging studies seem to suggest that a site for this ability may be a region of the left medial frontal cortex [Brodmann's area (BA) 8/9]. A number of studies using a variety of tasks have demonstrated activation in this region (Fletcher *et al*. 1995; Goel *et al*. 1995; Castelli *et al*. 2000; Gallagher *et al*. 2000; Vogeley *et al*. 2001). This region has also been implicated in mental state attribution in the only event-related potential study of ToM (Sabbagh and Taylor 2000). Other areas reported include regions of the left temporal lobe (Goel *et al*. 1995; Castelli *et al*. 2000; Gallagher *et al*. 2000) and

the temporo-parietal junction (Fletcher *et al*. 1995; Castelli *et al*. 2000; Gallagher *et al*. 2000).

As with any comparisons across neuroimaging studies, it is difficult to compare activations in different brain regions, due to the widely differing task demands in each of the studies. However, the common area that appears to consistently be activated in these types of studies in healthy individuals seems to be the left medial prefrontal cortex.

NEUROIMAGING STUDIES OF ToM IN AUTISM

Two studies have been reported exploring the neural correlates of ToM in individuals with autism or Asperger's syndrome (Happé *et al*. 1996; Baron-Cohen *et al*. 1999).

Happé *et al*. (1996) were the first to report findings from a PET study of five patients with Asperger's syndrome. Subjects in this study completed a story comprehension task requiring either an understanding of mental states (the ToM condition), physical states (the control condition) or unconnected sentences (a second control condition). Findings from the Asperger's and control brains were broadly similar and replicated to some degree findings from a previous study (with healthy controls only; Fletcher *et al*. 1995) with activation seen in the left medial frontal lobe (BA8 and BA9). In the comparison between groups for the ToM and physical stories, there was one region active in the patient but not the control group; this was also in the left medial frontal lobe, but located more ventrally, in BA9/10. This was in the context of a significantly lower accuracy score in the ToM condition. The authors suggest that activation of this region may reflect alternative strategies in the Asperger's group for solving social problems. Although the exact cognitive processes ascribed to BA9 and BA10 have yet to be fully described, it is suggested that this region may be involved in general problem-solving tasks. It is suggested from this work that the Asperger's group may be utilizing a more cognitive strategy to solve a social problem, whereas the control subjects may have a dedicated region of BA8 for this purpose.

These findings, together with those that have come from the face-processing literature, seem to suggest that, in autistic individuals and those with Asperger's syndrome, the normal brain specialisation (in the case of face processing) and perhaps circuitry (in the case of ToM) has failed to develop along a normal course. Instead, in these individuals, alternative brain regions are called upon to solve the same problems (recognise faces or solve ToM problems). This may be in the context of intact performance or impaired performance.

Brothers (1990) has proposed a network of neural regions which she calls 'the social brain'; the orbito-frontal cortex, the superior temporal gyrus and the amygdala. Baron-Cohen and colleagues (1999) tested the integrity of the

network in a group of high-functioning individuals with autism or Asperger's syndrome, using fMRI and a task designed to probe these regions. Baron-Cohen's 'eyes task' requires subjects to view an eye pair and choose (from two words) a word which best describes what those eyes are thinking or feeling, in a forced-choice paradigm (the ToM condition). The control condition used in this AB block design fMRI experiment was to view the same eye pairs but to make a gender decision (male or female). Six subjects with autism and 12 healthy controls (six male and six female) were included in the experiment. For both the gender decision and the ToM conditions, the control subjects were significantly more accurate. In keeping with Brothers' ideas, regions of the temporal lobe and frontal regions were activated by this study; however, it was the left dorsolateral prefrontal region (rather than orbitofrontal cortex) which was active. In the comparison between the controls and autistic individuals, there was significantly greater response by the control group in the left amygdala, while autistic individuals did not activate the amygdala at all. Previous studies have posited a role for the amygdala in the perception of emotions, particularly fear (Adolphs *et al*. 1995; Morris *et al*. 1996), and it may be that this region is also involved in cognition related to attributing the mental states of others. Left inferior frontal gyrus (corresponding to BA44/45) and left insula were also significantly reduced in the autism group. Baron-Cohen and colleagues suggest that amygdala dysfunction in these individuals may be one aspect of abnormal brain function that contributes to impairments in social functioning.

NEUROIMAGING STUDIES OF ToM IN SCHIZOPHRENIA

To date there is only one study which has examined ToM in patients with schizophrenia (Russell *et al*. 2000). They used fMRI and Baron-Cohen's 'eyes task' (described above) to explore abnormalities at the neural level in mental state attribution in this patient group. Five individuals with schizophrenia were compared with seven matched controls.

The neurocognitive network for normal mental state attribution (MSA) comprised middle, inferior frontal and middle temporal regions. The normative network found for mental state attribution was in line with previous findings (Baron-Cohen *et al*. 1999). In the context of poor performance on this task, it was shown that patients with schizophrenia had significant reductions in neural activity in the left hemisphere and in the region of the middle and superior frontal lobes, bordering the insula. The reductions in blood oxygen level-dependent response was specifically in the left inferior frontal gyrus (corresponding to BA44/45) relative to control subjects. The frontal under-activation in schizophrenia echoes functional and structural findings indicative of a frontal dysfunction in this patient population (Weinberger *et al*. 1994; Harrison 1999). A drawback of this study is that the patients were not grouped

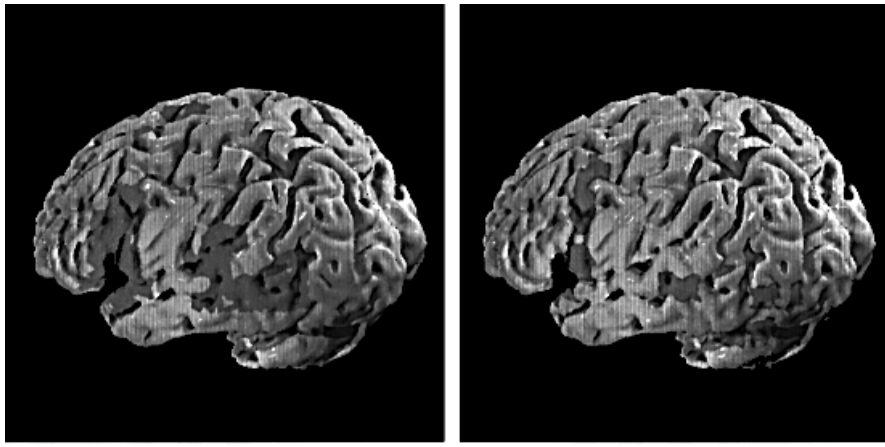


Figure 12.1. Generic brain activation maps for control group (a) and patients with schizophrenia (b): view of the left hemisphere. Voxel-wise probability of Type 1 error = 0.0004. Reproduced with permission from the American Psychiatric Publishing Inc., Russell *et al.* (2000)

on the basis of predominant symptomatology. The cost of scanning time can be prohibitive to this type of symptom-based approach, where large numbers of subjects are required. Future investigations would need to take into account the specific symptoms of the patients scanned. This was, however, the first demonstration of hypofrontality in schizophrenia during social processing.

The findings of Russell *et al.* (Figure 12.1) are consistent with Deakin's (1994) hypothesis of a dysfunction in schizophrenia in a basolateral circuit, involving ventral frontal and anterior temporal regions, leading to disturbances of social communication and interpretation. A dysfunction of the frontal part of this network in schizophrenia seems to be sufficient to cause the social interpretation deficits observed here during the MSA task. A reduction in a strikingly similar left inferior frontal focus has also been shown in patients with autism, known to show ToM deficits. The integrity of left prefrontal cortex seems thus to be crucial for intact ToM processing, as has been previously suggested (Goel *et al.* 1995; Frith 1996). A left frontal underactivation in schizophrenia during MSA thus confirms the hypothesis of a socio-emotional neurocognitive deficit in this group. Unlike the study reported by Baron-Cohen *et al.* (1999), there was no difference between the two groups in the region of the amygdala.

NEUROIMAGING STUDIES OF ToM IN PSYCHOPATHS

To date there have been no neuroimaging studies reported concerning ToM abilities in psychopaths. It is known that they do not differ from non-

psychopathic control subjects on ToM tasks (Blair *et al*. 1996); however, abnormalities in the limbic system of these individuals has been demonstrated on emotional tasks (Kiehl *et al*. 2001). A greater amount of frontotemporal activation in response to emotional words has also been demonstrated in this group (Intrator *et al*. 1997).

Given that the frontal lobes are known to be involved in ToM tasks, and that prefrontal grey matter volume is reduced in this group of individuals (Raine *et al*. 2000), it might be predicted that different neural pathways may be used by these individuals to solve ToM problems.

Emotional difficulties such as impaired fear processing and defective empathic capacity are the most instrumental and distinctive features of psychopathy. The deficiencies in empathy also compliment current research concerning the inability of psychopaths to regulate behaviour, due to difficulties in perceiving defenceless/surrender cues. The overriding emphasis in the past has been on the amygdala, as the most important biological structure of psychopathy. However, the orbitofrontal cortex may have a crucial role in emotional response in psychopaths. The whole of the prefrontal cortex seems to have a more complicated synergistic interaction in psychopathy. Davidson, Putnam and Larson (2000) provide a framework for a more dynamic interplay (or a lack of it) between neural regions that may explain psychopathy. It is known that the orbitofrontal cortex constrains impulsive outbursts, the amygdala pertains to fear and emotion, and the anterior cingulate cortex activates other regions of the brain in response to conflict. It is likely that, in psychopathic individuals, all of these structures play a role. The output from the amygdala to the hindbrain intensifies fear, whereas output to the hypothalamus serves to control autonomic fear responses (LeDoux 1998). In normal subjects, pleasant slides reduce the amplitude of the startle response elicited by abrupt noise-bursts, whereas aversive slides potentiate it (Vrana, Spence and Lang 1988). In people with psychopathy, a significant response attenuation is seen to aversive slides when compared to the neutral ones (Patrick, Bradley and Lang 1993). Thus, psychopathic individuals may not only suffer from a deficient fear and empathy responding but may also find expressions of fear and sadness in other human beings pleasurable.

SUMMARY — NEUROIMAGING STUDIES OF ToM IN PSYCHIATRIC DISORDERS

Imaging studies of ToM in schizophrenia and other psychiatric disorders are still in their infancy and the small sample sizes, as well as different paradigms used by different groups, make comparisons very difficult. While the scanning acquisition, parameters and methodological design of the tasks described above vary greatly (as do the results), some trends in this disparate data appear to be emerging. The most promising candidate for ToM is a region of the left

medial frontal gyrus (roughly BA8/BA9; Fletcher *et al*. 1995; Goel *et al*. 1995; Happé *et al*. 1996; Gallagher *et al*. 2000). This has been reported in a number of studies, although several authors also report more ventral regions of prefrontal cortex (BA44/45; Baron-Cohen *et al*. 1999; Russell *et al*. 2000). The imaging data seems to be at odds with the findings of lesion studies, which largely implicate the right hemisphere (Siegal, Carrington and Radel 1996; Winner *et al*. 1998; Happé *et al*. 1999; but see Rowe *et al*. 2001). These discrepancies have yet to be explained, although both methods of investigation (lesion and imaging studies) are limited by the small numbers of subjects used and the wide variation in task demands. One conclusion that can be drawn from these data is that one circumscribed brain area for ToM is unlikely to exist. The more likely explanation is that of a distributed network involving several brain regions working together. Further work in this area requires well-validated tasks that have extensive 'off-line' data, with a view to being suitable for functional imaging. In schizophrenia, it will be important to examine the effects of social cognition longitudinally in patients, as well as the effects (if any) of the newer generation of antipsychotic medications that are available today.

CONSIDERATIONS

As with any type of investigation, there are a number of considerations that should be borne in mind when both designing and executing a neuroimaging study and when reading about such investigations. These considerations include issues that are both clinical and methodological in nature, and some brief ideas relating to these are outlined below.

MEDICATION

As with any study using clinical populations, the issue of medication can be a thorny one. While it is possible in some studies, e.g. those looking at children with attention deficit hyperactivity disorder (ADHD) to take subjects off medication prior to scanning, this is rarely an option when dealing with adult psychiatric populations, e.g. patients with schizophrenia. Given that the majority of individuals in such studies will be medicated, and most likely on different types of medication, this point needs careful consideration. Particularly as it has been demonstrated that different types of antipsychotic medications may affect the blood oxygen level dependent response in different ways (Honey *et al*., 1999). Several suggestions can be made, however, for ways to minimise the impact of medication on the final results, e.g. studies of first break antipsychotic-naive subjects (which are inevitably difficult to find) or those on a particular type or class of medication, e.g. atypical antipsychotics or a particular kind of depot injection. Lastly, it may be possible within the

analysis to co-vary for medication dose (expressed either as a percentage of maximum dose or in chlorpromazine equivalents).

HOMOGENEITY OF THE SAMPLE

Issues relating to homogeneity of the sample are a particular problem in neuroimaging studies of schizophrenia, although increasingly investigations are focusing on more specialised questions relating to particular symptoms. While it may not be possible to scan the numbers of subjects required to explore all the major symptom subtypes (e.g. remitted, paranoid, non-paranoid, chronic deficit), studies that focus on just one group will yield findings that are more interpretable than those including any and all types of schizophrenic individuals. In cases where a range of symptoms are represented by the sample, it may be possible at the analysis stage to co-vary for symptom scores (or particular symptom subscale scores) as another strategy to remove variance related to symptom differences, rather than cognitive performance.

MATCHING FOR PERFORMANCE

Interpretation of differences in brain activity between two groups can be made difficult when one group is performing the task at a level that is significantly lower than another. This is often the case in studies with patients with schizophrenia, who will perform at a lower level relative to controls on a majority of tasks. What does it mean if groups of subjects are unable to do a task and show less brain activity? Would they still show less brain activity if they could do the task? There are a number of ways to combat this problem. First, there may be a subgroup of individuals who perform adequately on the task, e.g. patients in remission from schizophrenia do not differ from controls on ToM tasks; those higher-functioning autistic individuals are also able to pass ToM tests. Comparisons between these types of groups and controls, and the inclusion of a group who cannot do the task, may be more informative. Second, it may be possible at the analysis stage to co-vary for the level of performance, or split clinical subjects into groups based on a threshold (e.g. above and below a median), although the often small numbers of subjects in such studies may preclude this approach.

SOCIAL FUNCTIONING

Tasks designed for use in the scanner tend to be extremely artificial in nature. While numerous strategies to make the stimuli more ecologically valid can be employed, e.g. moving or three-dimensional faces, these types of design do not always fit easily in the parameters demanded by the scanning acquisition or analysis. One solution may be to include the more ecologically valid tests in a

pre- or post-scanning session and look at how performance on these relates to performance on the task completed in the scanner.

Similarly, the question of how performance and neural activity on a particular task relates to real-world functioning needs to be considered. In order to determine the relationship of the task in question to real-world social functioning, one strategy would be to include some measure of social functioning in the post- or pre-scanning battery, since social functioning is the Holy Grail of outcome measures in psychiatry.

CONCLUSIONS

Social cognition is a separate domain of cognitive function and is linked strongly to functional outcome. If we can find the neural basis of deficits of social cognition in psychiatric disorders, we may be able to find ways of remedying this deficit. It is certain that the neuroimaging data on social cognition, specifically ToM and face processing in clinical populations, is sparse. With relatively few studies (and within these relatively small numbers of subjects), conclusions at present can only be tentative. There are, however, some strands that seem to come together to provide some insights into the neural basis of social cognition in psychiatric disorders. Across both domains (face processing and ToM) the integrity of the amgydala appears to be crucial. Additionally, the frontal lobes (both orbitofrontal and medial frontal regions) seem to be important for ToM.

Social cognition might be viewed as a complex two-stage process with different neuronal circuits complementing each other. In day-to-day life we would look at the face of a person and then use that information to think further about his/her mental state. The emotion-processing aspect of this act is automatic, but the ToM component, where one is attributing a mental state to a person, is more cognitive and inferential.

Neuroimaging studies also show us that it is necessary for the subcortical and frontal systems to work in tandem. Subcortical limbic structures (e.g. amygdala) connect to the orbitofrontal cortex and information is transferred to the medial frontal lobe to make ToM decisions. In autism it is likely that the connections between the subcortical and the frontal cortical areas are not fully developed. Various studies have now shown a lack of activation in the fusiform gyrus (a region consistently activated in normal subjects in response to faces) and the left amygdala, perhaps due to structural deficits in these areas. Neuronal plasticity is now a recognised feature of the adult brain and it is not yet known whether these areas would light up with cognitive remediation techniques. In schizophrenia, there is evidence of abnormalities in brain activation for both face processing and ToM tasks, but there are many more confounding factors in this illness than in autism, as listed above. One of the

important aspects of future investigations might be the possible remediation of social cognition with the newer-generation antipsychotics, cognitive enhancers and cognitive remediation techniques. Lastly, psychopathic disorder provides us with a unique instance in which the abilities of social cognition seem to be intact but the cortical networks that are activated are different from normal controls. Social cognition is what makes us human. We would like to end this chapter with a quote from Steven Pinker, the famous psycholinguist, who has spent years researching language. He writes, in his book *How the Mind Works*:

> We mortals cannot read other people's minds directly. But we make good guesses from what they say, what we read between the lines, what they show in their faces and eyes, and what best explains their behavior. It is our species' most remarkable talent.

REFERENCES

Anskiewicz, A.S. (1979). Autonomic components of vicarious conditioning and psychopathy. *J Clin. Psychol.* **35**, 60–67.

Adolphs, R., Tranel, D., Damasio, H. and Damasio, A.R. (1995). Fear and the human amygdala. *J. Neurosci.* **15**, 5879–5891.

APA (American Psychiatric Association) (1994). *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV 4th edn*. American Psychiatric Association, Washington, DC.

Baron-Cohen, S. (1994). How to build a baby that can read minds. *Cahiers Psychol. Cogn.* **13**(5), 513–552.

Baron-Cohen, S. (1996). *Mindblindness: An Essay on Autism and Theory of Mind*. MIT Press, Cambridge, MA.

Baron-Cohen, S., Ring, H., Wheelwright, S., Bullmore, E. *et al.* (1999). Social intelligence in the normal and autistic brain: an fMRI study. *Eur. J. Neurosci.* **11**, 1891–1898.

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

Bellack, A., Blanchard, J. and Mueser, K. (1996). Cue availability and affect perception in schizophrenia. *Schizophr. Bull.* **22**, 535–544.

Blair, R.J. and Cipolotti, L. (2000). Impaired social response reversal. A case of 'acquired sociopathy'. *Brain* **123**, 1122–1141.

Blair, R.J., Jones, L., Clark, F. and Smith, M. (1997). The psychopathic individual: a lack of responsiveness to distress cues? *Psychophysiology* **34**, 192–198.

Blair, J., Sellars, C., Strickland, I., Clark, F. *et al.* (1996). Theory of mind in the psychopath. *J. Forens. Psychiat.* **7**, 15–25.

Bormann-Kischkel, C., Vilsmeier, M. and Baude, B. (1995). The development of emotional concepts in autism. *J. Child Adolesc. Psychiat.* **36**, 1243–1259.

Brekke, J., DeBonis, J.A. and Graham, J.W. (1994). A latent structure analysis of the positive and negative symptoms in schizophrenia. *Compr. Psychiat.* **35**, 252–259.

Brothers, L. (1990). The social brain: a project for integrating primate behaviour and neurophysiology in a new domain. *Concepts Neurosci* **1**, 27–51.

Castelli, F., Happé, F., Frith, U. and Frith, C. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage* **12**, 315–325.

Cheney, D., Seyfarth, R. and Smuts, B. (1986). Social relationships and social cognition in non-human primates. *Science* **234**, 1361–1366.

Corcoran, R., Cahill, C. and Frith, C.D. (1997). The appreciation of visual jokes in people with schizophrenia: a study of 'mentalizing' ability. *Schizophr. Res.* **24**, 319–327.

Corcoran, R., Mercer, G. and Frith, C.D. (1995). Schizophrenia, symptomatology and social inference: investigating 'theory of mind' in people with schizophrenia. *Schizophr. Res.* **17**, 5–13.

Critchley, H., Daly, E., Bullmore, E., Williams, S. *et al.* (2000). The functional neuroanatomy of social behaviour: changes in cerebral blood flow when people with autistic disorder process facial expressions. *Brain* **123**, 2203–2212.

Davies, S., Bishop, D., Manstead, A.S. and Tantam, D. (1994). Face perception in children with autism and Asperger's syndrome. *J. Child Psychol. Psychiat.* **35**, 1033–1057.

Day, R. and Wong, S. (1996). Anomalous perceptual asymmetries for negative emotional stimuli in the psychopath. *J. Abnorm. Psychol.* **105**, 648–652.

Davidson, R.J., Putnam, K.M., Larson, C.L. (2000). Dysfunction in the neural circuitry of emotion regulation — a possible prelude to violence. *Science* **289** (5479), 591–594.

Deakin, J.F.W. (1994). Neuropsychological implications of brain changes in schizophrenia. *Psychopathology* **27**, 251–254.

DeJong, A., Giel, R., Slooff, C. and Wiersma, D. (1986). Relationship between symtomatology and social disability. *Soc. Psychiat.* **21**, 200–205.

Doody, G., Gotz, M., Johnstone, E., Frith, C., and Cunningham-Owens, G. (1998). Theory of mind and psychosis. *Psychol. Med.* **28**, 397–405.

Drury, V., Robinson, E. and Birchwood, M. (1998). 'Theory of mind' skills during an acute episode of psychosis and following recovery. *Psychol. Med.* **28**, 1101–1112.

Edwards, J., Pattison, P., Jackson, H. and Wales, R. (2001). Facial affect and affective prosody recognition in first-episode schizophrenia. *Schizophr. Res.* **48**, 235–253.

Eslinger, P. and Damasio, A. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation: patient EVR. *Neurology* **35**, 1731–1741.

Feinberg, T.E., Rifkin, A., Schaffer, C. and Walker, E. (1986). Facial discrimination and emotional recognition in schizophrenia and affective disorders. *Arch. Gen. Psychiat.* **43**, 276–279.

Fiske, S.T. and Taylor, S.E. (1991). *Social Cognition*, 2nd edn. International Editions Psychology Series, McGraw-Hill, New York.

Fletcher, P., Happé, F., Frith, U., Baker, S. *et al.* (1995). Other minds in the brain: a functional imaging study of 'theory of mind' in story comprehension. *Cognition* **57**, 109–128.

Frith, C.D. (1996). Brain mechanisms for having a 'theory of mind'. *J. Psychopharmacol.* **19**, 9–15.

Frith, C.D. and Corcoran, R. (1996). Exploring 'theory of mind' in people with schizophrenia. *Psychol. Med.* **26**, 521–530.

Gallagher, H., Happé, F., Brunswick, N., Fletcher, P. *et al.* (2000). Reading the mind in cartoons and stories: an fMRI study of theory of mind in verbal and non-verbal tasks. *Neuropsychologia* **38**, 11–21.

Gardner, H. (1983). *Frames of Mind: the Theory of Multiple Intelligences*. Basic Books, New York.

Gessler, S., Cutting, J., Frith, C.D. and Weinman, J. (1989). Schizophrenics' ability to judge facial emotion: a controlled study. *Br. J. Clin. Psychol.* **28**, 19–39.

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

Green, M., Kern, R., Robertson, M., Sergi, M. and Kee, K. (2000). Relevance of neurocognitive deficits for functional outcome in schizophrenia. In T. Sharma and P. Harvey (eds), *Cognition in Schizophrenia: Impairments, Importance, and Treatment Strategies*. Oxford University Press, Oxford.

Happé, F., Brownell, H. and Wimmer, E. (1999). Acquired 'theory of mind' impairments following stroke. *Cognition* **70**, 211–240.

Happé, F., Ehlers, S., Fletcher, P., Frith, U. *et al.* (1996). 'Theory of mind' in the brain. Evidence from a PET scan study of Asperger syndrome. *NeuroReport* **8**, 197–201.

Harrison, P. (1999). The neuropathology of schizophrenia. A critical review of the data and their interpretation. *Brain* **122**, 593–624.

Hobson, R.P. (1986). The autistic child's appraisal of expressions of emotion. *J. Child Psychol. Psychiat.* **27**, 321–342.

Hobson, R.P., Ouston, J., and Lee, A. (1988). What's in a face? The case of autism. *Br. J. Psychol.* **79**, 441–453.

Hoffman, E. and Haxby, J. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neurosci.* **3**, 80–84.

Honey, G.D., Bullmore, E.T., Soni, W., Varatheesan, M., Williams, S.C.R., Sharma, T. (1999). Differences in frontal cortical activation by a working memory task following substitution of risperidone for typical antipsychotic drugs in patients with schizophrenia. *Proc. Natl Acad. Sci.* **96**, 13432–13437.

Intrator, J., Hare, R., Stritzke, P., Brichtswein, K. *et al.* (1997). A brain imaging (single photon emission computerized tomography) study of semantic and affective processing in psychopaths. *Biol. Psychiat.* **42**, 96–103.

Kiehl, K.A., Smith, A.M., Hare, R.D., Mendrek, A. *et al.* (2001). Limbic abnormalities in affective processing by criminal psychopaths as revealed by functional magnetic resonance imaging. *Biol. Psychiat.* **50**, 677–684.

Kohler, C., Bilker, W., Hagendoorn, M., Gur, R. and Gur, R. (2000). Emotion recognition deficit in schizophrenia: association with symptomatology and cognition. *Biol. Psychiat.* **48**, 127–136.

Langdell, T. (1978). Recognition of faces: an approach to the study of autism. *J. Child Psychol. Psychiat.* **19**, 255–268.

LeDoux, J. (1998). *The Emotional Brain*. Weidenfeld, London.

Lewis, S. and Garver, D. (1995). Treatment and diagnostic subtype in facial affect recognition in schizophrenia. *J. Psychiat. Res.* **29**, 5–11.

Liddle, P. (1987). The symptoms of chronic schizophrenia: a re-examination of the positive–negative dichotomy. *Br. J. Psychiat.* **151**, 145–151.

Mandal, M., Pandey, R. and Prasad, A. (1998). Facial expressions of emotion and schizophrenia: a review. *Schizophr. Bull.* **24**, 399–412.

Mitchell, P. (1997). *Introduction to Theory of Mind. Children, Autism and Apes*. Arnold, London.

Morris, V.J.S., Frith, C.D., Perrett, D.I., Rowland, D. *et al.* (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature* **383**, 812–815.

Mueser, K.T. and Bellack, A.S. (1998). Social skills and social functioning. In K. Mueser and N. Tarrier (eds), *Handbook of Social Functioning in Schizophrenia*. Allyn and Bacon, Boston, MA.

Novic, J., Luchins, D. and Perline, R. (1984). Facial affect recognition in schizophrenia. Is there a differential deficit? *Br. J. Psychiat.* **144**, 533–537.

Ostrom, T. (1984). The sovereignty of social cognition. In R.S. Wyer and T.K. Srull (eds) *Handbook of Social Cognition*, vol. 1. Erlbaum, New York, pp. 1–37.

Patrick, C.J., Bradley, M.M., and Lang, P.J. (1993). Emotion in the criminal psychopath: startle reflex modulation. *J. Abnorm. Psychol.* **102**, 82–92.

Penn, D. and Corrigan, P. (1997). Introduction: factors underlying social functioning in schizophrenia: information processing and social perception. *Psychiatry* **60**, 279–291.

Perrett, D. and Emery, N. (1994). Understanding the intentions of others from visual signals: neurophysiological evidence. *Cahiers Psychol. Cogn.* **13**, 683–694.

Perrett, D., Harries, M., Bevan, R., Thomas, S. *et al.* (1989). Frameworks of analysis for the neural representation of animate objects and actions. *J. Exp. Biol.* **146**, 87–113.

Perrett, D., Harries, M., Benson, P., Chitty, A. and Mistlin, A. (1990a). Retrieval of structure from rigid and biological motion: an analysis of the visual responses of neurons in the macaque temporal cortex. In T. Troscianko and A. Blake (eds) *AI and the Eye*. Wiley, Chichester, pp. 181–200.

Perrett, D., Harries, M., Mistlin, A. and Chitty, A. (1990b). Three stages in the classification of body movements by visual neurones. In H.B. Barlow, C. Blakemore and M. Weston-Smith (eds), *Images and Understanding*. Cambridge University Press, London, pp. 94–107, 337–378.

Perrett, D., Hietanen, J., Oram, M. and Benson, P. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Phil. Trans. R. Soc. Lond. Ser. B* **335**, 23–30.

Perrett, D., Mistlin, A. and Chitty, A. (1987). Visual cells responsive to faces. *Trends Neurosci.* **10**, 358–364.

Phan, K.L., Wager, T., Taylor, S.F., and Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage* **16**, 331–348.

Pickup, G. and Frith, C.D. (2001). Theory of mind impairments in schizophrenia: symptomatology, severity and specificity. *Psychol. Med.* **31**, 207–220.

Pierce, K, Muller, R., Ambrose, J., Allen, G. and Courchesne, E. (2001). Face processing occurs outside the fusiform face area in autism: evidience from functional MRI. *Brain* **124**, 2059–2073.

Pinker, S. (1997). *How the mind works*. W.W. Norton & Co., New York.

Phillips, M.L., Williams, L., Senior, C., Bullmore, E.T. *et al.* (1999). Investigation of the neural correlates of threat perception in paranoid schizophrenia: an fMRI study. *Psychiat. Res. Neuroimage.* **92***, 11–31.*

Prudo, R. and Monrow Blum, H. (1987). Five-year outcome and prognosis in schizophrenia: a report from the London field centre of the international pilot study of schizophrenia. *Br. J. Psychiat.* **150**, 345–354.

Raine, A., Lencz, T., Bihrle, S., LaCasse, L. and Colletti, P. (2000). Reduced prefrontal gray matter volume and reduced autonomic activity in antisocial personality disorder. *Arch. Gen. Psychiat.* **57**, 119–127.

Rowe, A., Bullock, P., Polkey, C. and Morris, R. (2001). 'Theory of mind' impairments and their relation to executive functioning following frontal lobe excisions. *Brain*, **124**, 600–616.

Russell, T.A. and Morris, R. (2001). Further independent evidence for symptom-specific deficits in theory of mind in schizophrenia. *Schizophr. Res.* **53** (suppl), 137.

Russell, T.A., Rubia, K., Bullmore, E., Soni, W. *et al.* (2000). Exploring the social brain in schizophrenia: left prefrontal underactivation during mental state attribution. *Am. J. Psychiat.* **157**, 2040–2042.

Sabbagh, M. and Taylor, M. (2000). Neural correlates of theory-of-mind reasoning: an event-related potential study. *Psychol. Sci.* **11**, 46–50.

Sarfati, Y., Hardy-Bayle, M.-C., Nadel, J., Chevalier, J.-F. and Widlöcher, D. (1997a). Attribution of mental states to others by schizophrenic patients. *Cogn. Neuropsychiat.* **2**, 1–17.

Sarfati, Y., Hardy-Bayle, M.-C., Besche, C. and Widlöcher, D. (1997b). Attribution of intentions to others in people with schizophrenia: a non-verbal exploration with comic strips. *Schizophr. Res.* **25**, 199–209.

Sarfati, Y., Hardy-Bayle, M.-C., Brunet, E. and Widlöcher, D. (1999). Investigating theory of mind in schizophrenia: influence of verbalization in disorganized and non-disorganized patients. *Schizophr. Res.* **37**, 183–190.

Schultz, R., Gauthier, I., Klin, A., Fulbright, R. *et al.* (2000). Abnormal ventral temporal cortical activity during face discrimination among individuals with autism and Asperger's syndrome. *Arch. Gen. Psychiat.* **57**, 331–340.

Schneider, F., Gur, R. and Shtasel, D. (1995). Emotional processing in schizophrenia: neurobehavioural probes in relation to psychopathology. *Schizophr. Res.* **17**, 67–75.

Schneider, F., Weiss, U., Kessler, C., Salloum, J.B. *et al.* (1998). Differential amygdala activation in schizophrenia during sadness. *Schizophr. Res.* **34**, 133–142.

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

Tantam, D., Monaghan, L., Nicholson, H. and Stirling, J. (1989). Autistic children's ability to interpret faces: a research note. *J. Child Psychol. Psychiat.* **30**, 623–630.

Vrana, S.R., Spence, E.L. and Lang, P.J. (1988). The startle probe response: a new measure of emotion? *J. Abnorm. Psychol.* **97**, 487–491.

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S. *et al.* (2000). Mind reading: neural mechanisms of theory of mind and self-perspective. *NeuroImage* **14**, 170–181.

Walker, E., Marwit, S. and Emory, E. (1996). A cross-sectional study of emotion recognition in schizophrenics. *J. Abnorm. Psychol.* **89**, 428–436.

Weinberger, D., Aloia, M., Goldberg, T. and Berman, K. (1994). The frontal lobes and schizophrenia. *J. Neuropsychiat.* **6**, 419–427.

Williamson, S., Harpur, T.J. and Hare, R.D. (1991). Abnormal processing of affective words by psychopaths. *Psychophysiology* **28**, 260–273.

Winner, W., Brownell, H., Happe, F., Blum, A. and Pincus, D. (1998). Distinguishing lies from jokes: theory of mind deficits and discourse interpretation in right hemisphere brain-damaged patients. *Brain Lang.* **62**, 89–106.

Wolwer, W., Streit, M., Polzer, U. and Gaebel, W. (1996). Facial affect recognition in the course of schizophrenia. *Eur. Arch. Clin. Neurosci.* **246**, 165–170.

13

Social Cognition and Behaviour in Schizophrenia

MARTIN BRÜNE

Centre for Psychiatry and Psychotherapy, University of Bochum, Germany

When Emil Kraepelin (1899) subsumed catatonia, hebephrenia, dementia paranoides, and (later) dementia simplex under the term 'dementia praecox', he proposed that these disorders were characterised by a severe cognitive decline in the course of the illness. Eugen Bleuler (1911), in contrast, not only emphasised a potentially more benign outcome, but also specifically rendered thought disorders, affective symptoms, ambivalence, autism and abnormalities in emotional expression and experience as core features of a cluster of disorders he labelled 'schizophrenia' (literally, 'split mind'). Many clinicians and researchers since then have felt uncomfortable with the traditional grouping of schizophrenia, mainly because of their uncertainty whether the schizophrenic subtypes represent reliable 'disease entities'. Modern classification systems, however, still subscribe to the original categorisation, although other conceptualisations, such as type I and type II schizophrenia (Crow 1985), or empirically-based subtyping according to behavioural phenotypes, have been proposed (e.g. Liddle 1987; Frith 1992). To date, however, it is still a matter of debate whether schizophrenia represents a set of heterogeneous disorders, whether a dimensional approach is more accurate, or whether the entire concept should be replaced by a symptom-orientated access (Penn *et al.* 1997).

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

This controversy notwithstanding, one of the most impressive common denominators of this group of disorders is a compromised social functioning of the affected individuals. Many studies have consequently endorsed that impaired social functioning be included in the diagnostic criteria of schizophrenia (DSM-IV; APA 1994). The underlying cognitive dysfunctions of social skill deficits in schizophrenia are, however, multifaceted and to some extent still obscure. Empirical research suggests that, among other factors, specific impairments of perception, processing and interpretation of socially relevant stimuli, referred to as 'social cognition', play a crucial role (Penn *et al*. 1997).

The present chapter, therefore, deals primarily with two cognitive key mechanisms involved in social interaction: emotion recognition and mental state attribution. The scientific angle from which these issues are addressed here proposes an essentially evolutionary stance, for a number of reasons. First, in contrast to traditional models of the social sciences, the view that natural and sexual selection has shaped human cognitive capacities through adaptation is more consistent with empirical studies of social reasoning (Cosmides and Tooby 1994). Psychological mechanisms evolved to solve specific problems of adaptive significance under ancestral conditions. For such an extraordinarily social species as human beings, many of these problem-solving psychological mechanisms and strategies likely emerged from social competition in terms of survival and reproduction. Second, our understanding of human psychological mechanisms is substantially supported by comparative neuropsychological and neurophysiological studies in non-human primates, suggesting a continuum of social cognition in primate evolution and its representation in a dedicated brain system (e.g. Brothers 1990; Whiten 2000; Brüne 2001, 2002; see also preceding chapters). Third, evolutionary theory represents the scientific basis for ethological observation, and this approach may provide essential cues for the understanding of the actual social behaviour of psychiatric populations (Troisi 1999). We begin, therefore, with a brief survey of the ethology of schizophrenia.

ETHOLOGICAL OBSERVATIONS IN SCHIZOPHRENIA

One of the first studies, based on naturalistic behavioural observation of chronic schizophrenic female inpatients over a period of 1.5 years (Staehelin 1953), revealed that the patients' behavioural repertoire was severely restricted to the defence of a personal 'territory', to maintaining a rigid social hierarchy, and to the avoidance of any body contact. Deliberate or accidental violation of a critical interpersonal distance, for instance, elicited stereotyped flight-or-fight responses. Staehelin was most impressed by the fact that he did not observe a

single case of friendship among the patients, neither did he recognise mutual help or emotions of mercy or empathy.[1]

More specifically, in a series of studies of non-verbal behaviour during interviews, Grant (1968) discovered that schizophrenic patients displayed behaviours related to ambivalence, submission and thwarted flight more frequently than psychiatric and non-psychiatric control groups. Other studies have confirmed that schizophrenic patients differ from patients with affective disorders in terms of their non-verbal behaviour on the ward, and that some findings of ethological examination may even have predictive value for the course and outcome of the disorder (e.g. Jones and Pansa 1979; McGuire and Polsky 1979, 1983). In addition, Pitman *et al.* (1987) found that the facial behaviour of non-medicated schizophrenic patients differed from controls, such that non-paranoid patients were impaired in their verbal expressiveness and showed only poor eye contact while talking, suggestive of flight behaviour, whereas paranoid patients had more eye contact but less eyebrow raising, which corresponds to staring as an agonistic behavioural correlate.

The study group of Krause and colleagues (1989) assessed the facial expression of schizophrenic patients engaged in conversation with lay persons who were unaware of the purpose of the study and even did not know that they were talking to patients. The theoretical background was basically psychoanalytic, proposing that persons with 'structural' disturbances, such as schizophrenia, were impaired in regulating intimacy and distance in social interactions. When comparing the facial expression of the schizophrenics and their interlocutors, using the Emotional Facial Action Coding System devised by Friesen and Ekman (1984), it turned out that the schizophrenic patients displayed fewer upper face activities, showed fewer primary emotions, such as happiness, anger, disgust, contempt, fear, surprise and interest, but more negative emotions than the healthy persons. In particular, action units indicating commitment in a conversation, e.g. raising or lowering the eyebrows, were less frequent in the schizophrenic group. The authors interpreted these findings in terms of a rejecting style of interaction in order to maintain greater distance from their interacting partners (Krause *et al.* 1989).

More recently, Troisi and his co-workers (1998) assessed drug-free patients with schizophrenia who were videotaped during interviews. The non-verbal behaviour was rated according to an ethological scoring system. Schizophrenic patients scored lower on pro-social behaviour, gesture and displacement activities than a control group of students who were videotaped during social stress (medical examination). This difference indicates a globally

[1] Note that the observed patients were non-medicated, since neuroleptics only became available in the early 1950s!

impaired non-verbal expressiveness and a deficit in encoding socially relevant signals among the patients. Interestingly, these findings were largely independent of psychopathological measures, i.e. negative and positive symptoms, which suggests that impaired social interaction reflects a separate dimension of schizophrenic disorders (overview in Troisi 1999).

These observation-driven studies therefore imply that poor social functioning in schizophrenia may be correlated with difficulties in understanding the emotional expressions and the intentions and dispositions of other individuals. Thus, it may be useful to take a closer look at the capacity to process emotional cues and at the way schizophrenics infer the mental states of others. We will start with the phylogenetically older system of emotion recognition.

EMOTION RECOGNITION IN SCHIZOPHRENIA

INTRODUCTORY REMARKS

Charles Darwin was probably the first to recognise the significance of emotional expressions for communication, and we may take for granted that the ability to perceive and to interpret emotional states of con-specifics has been crucial for all primates and many other animals, including man (Darwin 1872). Many such social signals in humans are encoded by a universally comprehensible set of displays of facial expression of emotions, although emotion recognition clearly also involves other modalities, such as gestures and vocal communication (e.g. Izard 1971; Ekman and Friesen 1976; Ploog 1999).

Numerous studies conducted since the late 1950s have shown a consistent impairment of affect recognition in schizophrenia (e.g. Izard 1959; Dougherty Bartlett and Izard 1974). Nevertheless, the extent to which emotion recognition is impaired in schizophrenia, and the relationship of such deficits to the actual social behaviour of schizophrenic subjects, is a rather complex question. These early studies, for instance, did not employ rigorous diagnostic criteria, neither did they refer to patient-specific, illness-related and other interfering cognitive variables. Thus, there has been considerable debate as to whether the deficit in affect recognition in schizophrenia is specific, or whether it represents a more general deficit of face processing, or whether it is even secondary to a global cognitive impairment. Moreover, the results are inconsistent regarding the relationship of affect recognition impairment to subject variables, e.g. sex, age, etc., and to illness variables, e.g. acuteness, duration and subtype of the disorder, number of hospitalisations, medication, and measures of psychopathology. Likewise, it is important to address whether affect recognition deficits are state- or trait-dependent and therefore possibly associated with other vulnerability markers, and whether they are at all related to the actual social competence and social skills of schizophrenic patients (e.g. Penn *et al.* 1997).

Despite the effort here to present a concise overview, one has to keep in mind that the studies of emotion recognition in schizophrenia differ substantially, e.g. regarding the mode of stimulus presentation. Most researchers have used standardised pictorial presentations of facial expressions of emotion, following Ekman and Friesen (1976) or Izard (1971). Furthermore, newer studies have included other measures of emotions, such as gestures, body posture and vocal cues, in order to mimic 'real-life' situations more accurately. Also, the test batteries for the assessment of other cognitive variables, psychopathology and social skills vary considerably. Hence, the comparability of studies is constrained by divergent study designs. Finally, it must be taken into account that the sample size of most studies is limited to roughly 20+ patients, such that subtyping within the sample provides little power to detect significant correlations (Poole, Tobias and Vinogradov 2000).

EMOTION RECOGNITION IN SCHIZOPHRENIA — A DIFFERENTIAL DEFICIT OR GENERAL COGNITIVE IMPAIRMENT?

Most studies suggest that subject variables such as age and sex have little influence on the performance of schizophrenics on emotion recognition tests (e.g. Schneider *et al.* 1995). Age and sex differences, however, have not been extensively addressed so far, and many studies are clearly biased towards male schizophrenic subjects (e.g. Bryson, Bell and Lysaker 1997; Sweet *et al.* 1998). Likewise, the poor emotion recognition abilities in schizophrenia do not seem to be directly related to antipsychotic medication (e.g. Cutting 1981; Salem, Kring and Kerr 1996; Wölwer *et al.* 1996). Nevertheless, this raises the question whether emotion recognition impairment in schizophrenia is associated with the acuteness or duration (i.e. chronicity) of the disorder, and to psychopathological measures or symptom clusters.

In one of the first studies applying Research Diagnostic Criteria (RDC) for schizophrenia (Spitzer, Endicott and Robins 1975), Cutting (1981) asked patients with acute and chronic schizophrenia to choose, from pairs of photographs, which one displayed the more friendly facial expression. This task was compared to age estimation and rating of 'meanness' of faces, as well as to a task involving the judgement of the 'typicalness' of colours. Acute, but not chronic, schizophrenic patients performed poorly on the friendliness decision task in comparison to controls. A difference was also found in rating the meanness of faces, whereas there was no such difference in judging age or colours. The performance on affect recognition was related to neither IQ nor medication. Novic, Luchins and Perline (1984) assessed facial affect recognition using standardised pictures (Izard 1971) and an auditory affect recognition task. In addition, a task involving the recognition of faces, in which the subjects had to choose a target photograph from six other photos, was applied. Contrary to previous studies, no difference between chronic schizophrenic

patients and healthy subjects was found, at least after eliminating items of low discriminatory power and after general face recognition performance was co-varied out. Moreover, there was no association with the severity of blunted affect, suggesting that impaired facial affect recognition in schizophrenia is neither a differential nor a specific deficit, nor is it related to psychopathology. Feinberg and his co-workers (1986) compared schizophrenic patients with depressed and normal controls in their abilities in emotion recognition and verbal labelling of facial expressions of emotion, according to Ekman and Friesen's (1976) classification of emotions (angry, happy, sad, fearful, disgusted, surprised and neutral faces). In addition, performance on tasks involving identity matching of two persons was assessed. The schizophrenic group performed significantly worse than the control groups on the emotion-discrimination task and, at a trend level, more poorly on the emotion-labelling task. Moreover, the study revealed that schizophrenic patients processed facial stimuli at a slower rate. The authors concluded that the results were consistent with a generalised deficit of processing facial stimuli in schizophrenia, but that a more specific deficit in facial affect recognition could not be ruled out. Cramer, Weegmann and O'Neill (1989) used audio-visual tapes comprising scenes of interactions between two characters. Patients and controls were asked to describe each scene's main emotional content, as chosen from a list of emotional states. The schizophrenic patients judged the scenes differently, often in the opposite direction, compared to controls. This finding was apparently related to the ambiguity of the scenes, because the difference between schizophrenic and control subjects in the two most unambiguous scenes was statistically not significant. No association with psychopathology, e.g. paranoid symptoms, formal thought disorder and flattened affect, or with duration of inpatient treatment, was found. By contrast, Lewis and Garver (1995) found trend-level associations of facial affect recognition with positive and negative symptoms. Task performance did not improve 2 weeks after neuroleptic treatment compared to the performance at baseline, although the BPRS score declined significantly. No correlation was found with age at onset of the disorder. This study revealed, however, that paranoid patients performed significantly better than non-paranoid patients. Recently, similar results were found in a study using a more complex measure of social perception, the Profile of Non-verbal Sensitivity Test (PONS), devised by Rosenthal *et al.* (1979). The PONS comprises a set of audio and/or video scenes depicting visual cues from the face or the body; vocal cues were altered, such that the content was obscure. The scenes illustrated either positive or negative affective valence, and dominance or submission, respectively. The schizophrenic patients' understanding of the social content of the scenes was impaired compared to the task performance of healthy subjects. Whereas there was no effect of age or medication, patients with paranoid schizophrenia performed better than subjects with undifferentiated schizophrenia. No

relation was found between task performance and positive or negative symptoms, but performance was significantly inversely correlated with disorganisation (Toomey *et al.* 2002).

In particular, it has been debated whether emotion recognition deficits in schizophrenia are specifically related to *social* cognition or rather to a global cognitive dysfunction. To date, however, abundant research has not clearly resolved this question. On the one hand, the 'non-specificity' section has emphasised the influence of other cognitive variables on patients' performance on emotion recognition tasks. In a series of experiments, for instance, the study group of Kerr and Neale employed a test battery based on the Ekman and Friesen photographs (except that the 'disgusted' facial expression of emotion was replaced by 'ashamed'), comprising an emotion identification task and a facial emotion discrimination task (Kerr and Neale 1993). For comparison, the Facial Recognition Test, introduced by Benton *et al.* (1978), was presented. The schizophrenic subjects performed more poorly than the controls on all three tasks, but the poorer performance was not specifically related to identification and discrimination of emotional expressions. These results point to a generalised deficit of face processing, rather than a specific deficit of facial affect recognition. There was no correlation with psychopathology scores, neuroleptic treatment or duration of hospitalisation, or with chronicity and severity of the disorder. Instead, a correlation occurred with the number of previous hospitalisations (Salem, Kring and Kerr 1996). Mueser and her co-workers, using the same test battery, confirmed the results of the Kerr and Neale group, except that they found the divergent result that the performance on the face-processing tasks was associated with the chronicity of the disorder and with negative symptoms. However, no difference was found between acute schizophrenic patients and controls (Bellack, Blanchard and Mueser 1996; overview in Mueser *et al.* 1997). Likewise, Kohler *et al.* (2000) found schizophrenics to perform more poorly than the control groups on emotion recognition and age recognition without a differential deficit. Subjects and controls were asked to judge happy, sad and neutral faces on a seven-point Likert scale, ranging from very happy to very sad, and to estimate the ages of depicted persons ranging from their teens to their 70s. Both patients and controls made more errors in identifying emotion in female faces than in male faces, and in identifying sad in comparison to happy faces. The emotion recognition deficit, but not the age recognition deficit, correlated with alogia, hallucinations and formal thought disorder, with attention, verbal and spatial memory, and with language abilities. This result therefore was not consistent with the specificity hypothesis of the emotion recognition deficit in schizophrenia. Similarly, Whittaker, Deakin and Tomenson (2001) studied face and non-face tasks under four categories, i.e. perception, recognition memory, naming, and executive functioning. It turned out that the performance on the Benton Test of Facial Recognition (Benton *et al.* 1978),

and on other visual–spatial abilities tests, appeared to be related to medication; whereas emotion recognition deficits were associated with other non-face naming tasks, and impairments in matching faces by either identity or emotion were associated with poor performance on the Wisconsin Card Sorting Test. These results indicate disease-related deficits of semantic retrieval and of executive function, which, in sum, contradicts the assumption of a specific face-processing deficit in schizophrenia.

On the other hand, several studies suggest the opposite, namely that emotion recognition is specifically impaired in at least some patients with schizophrenia when compared to healthy controls and to other psychiatric populations. Bryson, Bell and Lysaker (1997), for instance, developed a new audiovisual discrimination task, including facial, voice-tonal and upper body cues, called the Bell–Lysaker Emotion Recognition Task (BLERT). The tapes consisted of 21 vignettes depicting the basic expressions of emotions according to Ekman's and Friesen's categorisations, including three standard monologues, which were paired with each of the seven emotional states. The performance of schizophrenic patients on the BLERT was correlated with measures of executive functioning, general intelligence, attention, memory, verbal learning, understanding of proverbs, and measures of psychopathology. In one study, the schizophrenic group was divided into good and bad performers on BLERT. The performance on BLERT correlated with attention, concentration, cognitive flexibility and short-term recognition memory, as well as with the ability to discriminate between relevant signals and unnecessary information. No correlation was found with global functioning, serial, visual or contextual memory, or thought disorder. However, in a regression analysis, the cognitive test variables accounted only for a relatively small proportion of variance, indicating that emotion recognition impairments in schizophrenia may represent a differential deficit that is only moderately influenced by other cognitive capacities (Bryson, Bell and Lysaker 1997). However, the authors hypothesised that, in comparison with studies applying still photographs, the BLERT may more complexly measure a subject's ability to judge the internal experience of another person according to behavioural signs, suggesting a link to theory of mind (ToM) abilities.

In a second study, a group of clinically stable schizophrenic patients performed more poorly than controls on both positive and negative emotion recognition tests, although negative affect recognition was more impaired than positive affect recognition. In addition, a greater response dispersion in the schizophrenic group was found, i.e. the schizophrenic subjects were more likely to misidentify a given affect with one of the other possible affects of choice. Interestingly, the schizophrenics confused affects that were not commonly mistaken by the controls, such as disgust with surprise or sadness, with neutral affect (Bell, Bryson and Lysaker 1997).

Since the impact of negative symptoms on facial affect recognition had been ambiguous in previous studies, Bryson *et al*. (1998) compared the performance of a subgroup of schizophrenic patients with a so-called 'deficit syndrome' to a non-deficit schizophrenic group. The deficit syndrome was characterised by a lack of social interest, affective flattening, a 'diminished sense of purpose', and symptom stability over a period of 1 year; whereas 'reactive symptoms', such as social withdrawal, were not included in the definition in order to distinguish the deficit syndrome from patients with negative symptoms. The deficit syndrome group performed significantly more poorly on the BLERT than the non-deficit group. The item 'diminished sense of purpose' was most strongly associated with affect recognition impairment in the deficit group; whereas IQ differences and the traditional negative symptomatology did not account for these findings. These results imply that patients with a deficit syndrome are more severely impaired relative to other schizophrenics in processing and responding to complex social stimuli.

In a similar experiment, Sweet *et al*. (1998) assessed schizophrenic patients with blunted affect in comparison with a group of schizophrenics without blunted affect, in terms of emotion recognition and emotional self-experience. A video technique to assess emotion perception was applied, in addition to self-rating of emotional experience and measures for attention and emotional blunting. The study revealed, however, that emotion recognition and emotional self-experience were largely independent of emotional blunting, suggesting separate neural systems for expression, perception and experience of emotional states. These results were confirmed in a subsequent study by Shaw *et al*. (1999). The assessment battery comprised a facial identity task, a facial affect discrimination task, a naming and selecting facial affect task, a task matching facial expressions of emotions, and ratings of emotional and non-emotional prosody matched with facial expression, and vice versa. In addition, an acoustic analysis of speech characteristics of the schizophrenic subjects was carried out. The schizophrenic subjects performed more poorly than controls on all tests. However, no association between impaired expression of emotion and the emotion recognition deficit was found. Interestingly, there was a correlation between inappropriate affect in the schizophrenic group and impaired facial affect recognition. The authors therefore speculated that the failure of schizophrenic patients to correctly recognise facial expression of emotions may interfere with their ability to attune their own affective states (Shaw *et al*. 1999).

In order to dissect the nature of emotion recognition in schizophrenic disorders at the level of perception and of unconscious processing, facial affect identification has been studied in a series of interesting experiments. In a study of visual scanning of faces, Streit, Wölwer and Gaebel (1997) found that, relative to controls, schizophrenic patients focus more on the regions between the eyes than on the eyes directly. Although visual exploration abnormalities

were associated with affective flattening, affective flattening was not correlated with facial affect recognition, neither was the performance on emotion recognition tasks associated with eye-movement abnormalities; thus, impaired facial affect recognition is not simply a function of abnormal visual scanning. Similar results were found in a Japanese sample of schizophrenic patients, a finding which interestingly suggests that such abnormalities of visual scanning in schizophrenia is probably largely independent of culture (Shimizu *et al.* 2000).

Interestingly, the use of facial expression of emotions as primes influences the judgement of subsequently presented neutral faces differentially in schizophrenia and controls. In a prime-mask task, where faces were presented in such a manner that the subjects were unable to become aware of the prime, schizophrenic patients and controls had to judge whether an emotionally neutral target face was pleasant or unpleasant. When positive or neutral primes were replaced by primes consisting of negative emotional facial expressions, the schizophrenics valued neutral target faces as significantly more unpleasant, whereas controls were less affected by the alteration in the primes. Negative symptoms were correlated with the magnitude of the negative judgement shift at trend level significance, and disorganised symptoms were inversely associated with the negative judgement shift. The authors concluded that schizophrenic patients may be less able to suppress emotionally relevant stimuli, possibly leading to a greater autonomic stress response, and that increased spreading of emotional information may be related to poor social functioning (Höschel and Irle 2001).

Moreover, Federman *et al.* (1998) found that laterality effects in facial affect recognition may play a role in schizophrenia. When subjects were presented with facial affect recognition and facial recognition tasks in addition to chimerical facial affect identification tasks, two groups emerged, according to right or left visual field biases. Schizophrenic subjects with a left visual field bias performed better in identifying sad faces than subjects with a right visual field bias. No such difference was found in identifying happy or angry faces. In healthy controls, happy and angry faces were identified more accurately by subjects with a left visual field bias. The fact that fewer schizophrenic patients had a left visual field bias than the control group could be indicative of a reduction of the normal right-hemisphere advantage for processing facial affect in a subset of schizophrenic patients (Federman *et al.* 1998).

EMOTION RECOGNITION DEFICIT — STATE- OR TRAIT-DEPENDENT?

Social cognitive impairments in schizophrenia, as described above, may fluctuate across different stages of the disorder. It has been argued, for instance, that the emotion recognition deficit in schizophrenia may represent a specific vulnerability or trait marker, which is likely to worsen during periods

of symptom exacerbation and during a chronically deteriorating course of the illness (Penn *et al*. 1997). To test for stability of the affect recognition deficit, Gaebel and Wölwer (1992) assessed the spontaneous facial expression of schizophrenic patients during a standardised interview, the performance on facial affect recognition tasks according to Ekman's and Friesen's (1976) basic classification of facial expressions of emotions, and the ability of schizophrenic subjects to imitate and simulate emotional states within 3 days after admission and 4 weeks after neuroleptic treatment. As expected, the schizophrenic group performed more poorly than healthy controls on facial affect recognition. More importantly, the deficit remained stable over time, despite improvement of psychotic symptoms. The authors noted also a stable deficit of spontaneous facial expressiveness, as well as a consistent impairment of imitation and simulation of emotional states in schizophrenia, suggesting trait-like deficits of facial affect recognition and of the involuntary encoding of facial expressions of emotions. No consistent association with psychopathology and medication was found. These findings were replicated in a subsequent study, in which a subgroup of acute schizophrenic patients with persisting affective flattening were assessed a third time, 8 weeks after admission (Wölwer *et al*. 1996). As in the previous study, the emotion recognition deficit remained almost unchanged. Similarly, in a study using a more complex BLERT but using a longer test–retest interval, stable emotion recognition deficits in schizophrenia were found over a 5 month period (Bell, Bryson and Lysaker 1997).

Most interesting for the question of whether emotion recognition in schizophrenia is a trait-like deficit, first-episode schizophrenic patients in partial remission were assessed using a computerised version of the Feinberg *et al*. (1986) study design (Edwards *et al*. 2001). In addition, affective prosody recognition was examined, compared to patients with other psychotic disorders and affective disorders and to non-psychiatric controls. The schizophrenic subjects and patients with other psychotic disorders were impaired in facial affect recognition, compared to subjects with affective psychoses and to healthy controls, in particular regarding recognition of negative emotions, such as fear and sadness. This finding was consistent across modalities, i.e. facial affect and affective prosody were equally involved. Psychopathology and IQ only had modest impact on task performance, in sum lending support to the hypothesis that, if schizophrenic subjects already have emotion recognition deficits during their first psychotic episode, the impaired capacity to recognise emotional states of others may even precede the onset of the disorder (Edwards *et al*. 2001).

EMOTION RECOGNITION DEFICIT AND SOCIAL COMPETENCE

An intriguing question remains as to whether emotion recognition deficits in schizophrenic patients are directly related to their actual social competence,

which intuitively should be the case. As noted above, however, perception, experience and expression of emotional states need not necessarily be linked, at least if these functions have separate underlying neural substrates (Sweet *et al.* 1998). Bellack and colleagues (1990) assessed the social skills of schizophrenic subjects with predominantly negative symptoms in comparison with non-negative schizophrenics, patients with affective disorders and healthy controls. The subjects were given 12 role-play tasks. The subjects' behaviour was assessed using video-tape recordings of gaze, speech duration, 'meshing' (smoothness of turn-taking and of conversational pauses), affect, verbal content, and their relationship to social adjustment and quality-of-life ratings as measures of community functioning. The main finding was that schizophrenic patients with negative symptoms performed most poorly on all measures, followed, in order, by the non-negative patients, patients with affective disorders, and normal controls. This result implies that negative symptoms may reflect impairments of non-verbal behaviour and affective attunement, and therefore have a potentially detrimental impact on interpersonal relations. In two studies, facial affect recognition and social competence of schizophrenic subjects were assessed using a conversation probe (Penn *et al.* 1995), similar to the method of Bellack *et al.* (1990), and behavioural observation. Mueser *et al.* (1996) found chronic schizophrenics to perform more poorly than controls on facial affect recognition tasks and on the Test of Facial Recognition (Benton *et al.* 1978), independent of medication. The emotion recognition deficit was weakly associated with social skills, as measured by the conversational probe, but more strongly with social adjustment ratings on the ward. These results were replicated in a different sample of chronic schizophrenic inpatients (Penn *et al.* 1996). In this study, affect recognition deficits were associated with reduced social competence, social interest and grooming, independent of other cognitive measures. Thus, impaired recognition of emotional states in others may be related to impaired self-perception (overview in Mueser *et al.* 1997; Penn *et al.* 1997). These findings, however, could not be replicated in a sample of schizophrenic outpatients, where a similar test battery was applied, including an additional videotape-based test for social cue recognition, according to which the subjects had to answer 'concrete' and 'abstract' questions about what happens in the depicted interactions between two or three characters (Ihnen *et al.* 1998). It turned out, contrary to expectations, that facial affect recognition and social cue recognition was only weakly correlated with social skills. Interestingly, self-ratings of the subjects' own social skills were highly associated with their performance on the conversational probe, even after ratings of psychopathology and demographics had been co-varied out. The authors hypothesised that the divergence between these results and previous studies may have been due to the brief time frame of the probe, or that emotion recognition and social competence may be less functionally related than previously assumed (Ihnen *et al.* 1998).

Recently, the study of Poole, Tobias and Vinogradov (2000) shed new light on the question of whether emotion recognition and social skills are interrelated. The study addressed the role of facial and vocal affect recognition in schizophrenia in relation to psychopathology, other cognitive variables, and social functioning in the community. There was some evidence that emotion recognition depends on semantic memory, abstract reasoning and executive-attentional capacities, but no correlation with age, sex or medication was found. Contrary to other studies, no association of emotion recognition with negative symptoms, excitement or depression–anxiety items emerged. However, facial and vocal affect recognition impairments were likewise associated with disorganised symptoms non-emotional cognitive symptoms and poor interpersonal relationships. These co-variates did not correlate with one another, suggesting that emotion recognition may have a mediating impact on these dimensions. The authors further speculated that a single neurocognitive defect may account for the association of poor social functioning and disorganised behaviour with affect recognition deficits in some patients (Poole, Tobias and Vinogradov 2000).

THEORY OF MIND IN SCHIZOPHRENIA

INTRODUCTORY REMARKS

In contrast to the phylogenetically (relatively) old ability to 'read' the emotional signals of other individuals — mediated by perception of social cues from facial expression, gestures, body posture and vocalisation — the capacity to infer what others believe, think, intend and pretend, without necessarily referring to direct sensory input, is characteristic of the evolution of hominids. Past and current empirical evidence of social inference abilities of the great apes and human infants suggests a continuity model of the evolution of metarepresentational cognitive capacities in hominids (reviewed by Suddendorf and Whiten 2001). A stimulating hypothesis about why this capacity evolved in primates is that it has been essential to cope with the demands of the social environment in terms of survival and reproductive success (e.g. Dunbar 1998). Several terms have been proposed for this ability, such as having a 'theory of mind (ToM)', 'mental state attribution', 'metarepresentation', taking the 'intentional stance', 'mentalising', and 'reflexive awareness' (overview in Langdon and Coltheart 1999). Developmental psychologists have dissected the ontogeny of ToM in many empirical studies (see e.g. Volkmar *et al.*, and Kain and Perner, in this volume). ToM involves the ability not only to represent what others may believe, but also to comprehend that one may hold false beliefs about the 'physical' and the 'mental' world. More sophisticated cognitive capacities related to ToM involve the understanding of metaphor,

irony and *faux pas*, because these processes involve the ability to go beyond the literal meaning of utterances in communications by inferring what the speaker actually might have intended. Whereas 3–4 year-old children are able to understand false beliefs, understanding metaphor and irony may only emerge around the age of 7, and *faux pas* even beyond 10 years of age. From a psychopathological perspective, abundant research exists about ToM deficits in autistic spectrum disorders and other developmental disorders (see Chapters 9 and 10).

For the purposes of this chapter, the crucial questions ask whether ToM deficits exist in schizophrenic disorders, and whether such impairments of mentalising may even account for a subset of psychotic symptoms. In other words, does schizophrenia represent a group of disorders in which a once 'normally' developed ToM deteriorated due to a disruption of the underlying cognitive systems?

Several theoretical models exist about the cognitive architecture of ToM abilities. These models cannot be outlined and discussed in detail here, but it is important to note that different theoretical frameworks may account for divergent interpretations of ToM studies in schizophrenia. The 'metarepresentational' theory ('theory–theory') proposed by Perner (1991) advocates a non-modular approach. It proposes that children develop primary representations in infancy that model reality, depending on what is directly perceived. They achieve secondary representation during the second year, allowing them to distinguish between reality (something held in view) and hypothetical situations (something held in mind). Eventually they accomplish metarepresentational capacities, which puts them into the position to apply general knowledge about representations to the understanding of other persons' representations, including possible misrepresentations, i.e. false beliefs (Perner 1991). The representational theory holds that the crucial differentiation between reality and mental models may not only underlie ToM but also the more general capacity to entertain multiple mental models simultaneously, and to 'collate' primary and secondary representations (Suddendorf and Whiten 2001).

Similarly, the 'simulation' theory suggests that mentalising is primarily related to the ability to imaginatively put oneself into the perspective of others (e.g. Davies and Stone 1995). This could imply that defective mentalising may result from one of two impairments: first, the incapacity to inhibit cognitively more salient, and thus distracting, information, instead of relying on less salient information (the 'disengagement hypothesis' according to Langdon and Coltheart 1999); second, 'an impaired ability to reason consequentially on the basis of hypothetical states' (the 'executive planning hypothesis', according to Langdon and Coltheart 1999).

A modular model of metarepresentation has been put forward by Leslie and colleagues (recently updated by Scholl and Leslie 1999). This model proposes

that cognitive capacities are domain-specifically organised in the brain, such that only a certain class of information is processed within a given module. A 'ToM mechanism' would therefore operate on a metarepresentational basis to infer mental states of other individuals. Scholl and Leslie (1999) have furthermore argued that the accurate functioning of the ToM mechanism interferes with a 'selection processor', which is necessary to segregate relevant and irrelevant information, such that the likelihood increases that an individual's inference of others' mental states is correct. All these models suggest that social cognition inherently bears the risk of error, i.e. of making false assumptions about the intentions of other individuals.

During the past decade, several scholars have proposed a ToM deficit in schizophrenia that may explain at least a subset of psychotic symptoms.[2] It is intuitively obvious, for example, that people who are unable to understand that beliefs are subjective representations of reality fail to distinguish between subjectivity and objectivity. The maintenance of delusional convictions may arise from this failure. Moreover, if individuals disregard the social signals of others and do not take into account their intentional stance, a breakdown of communication (and possibly the emergence of positive thought disorder) would consequentially emerge. Likewise, people who are impaired in monitoring their actions as consequences of their own intentions may be at risk of developing delusions of alien control (Frith 1992; overview in Langdon and Coltheart 1999).

In his original formulation, based on the modular metarepresentational approach, Frith (1992) has therefore argued that impaired mentalising abilities in schizophrenia may relate to: (a) disorders of willed action, including negative and disorganised symptoms; (b) disorders of self-monitoring, including delusions of alien control and voice-commenting hallucinations; and (c) disorders of monitoring other persons' thoughts and intentions, including delusions of reference and persecution. In reverse, it has been proposed that the mentalising abilities of schizophrenic patients could be predicted on the basis of their symptomatology such that patients with negative or positive behavioural symptoms would perform worst on ToM tasks, as would subjects with autism, due to their incapacity to represent mental states at all. Patients with paranoid symptoms were expected to perform more poorly than controls, due to their difficulties in monitoring other people's intentions. Patients with passivity symptoms and patients in remission were predicted to perform normally on ToM tasks (Pickup and Frith 2001).

By contrast, Hardy-Baylé (1994) has hypothesised that impaired mentalising abilities in schizophrenia are primarily related to a deficit in action-planning,

[2] Much earlier, however, Klaus Conrad (1979/1958) argued from a Gestalt-theoretical point of view that impaired perspective taking and a compromised self-reflection are core symptoms of the beginning of schizophrenia.

such that the absence of a mental representation of this action would lead to the incapacity to attribute mental states to the actions of others. She thus advocates a non-modular model of executive planning deficits in schizophrenia, accounting for a compromised ToM. Consequently, such mentalising impairments would be expected to be most prevalent in schizophrenic patients with disorganised symptoms, as opposed to those with non-disorganised symptoms. Moreover, Hardy-Baylé (1994) has proposed independent deficits in understanding 'intentions' and 'beliefs', since intentions and beliefs have been supposed to involve different cognitive processes.

On the contrary, Walston *et al.* (2000) have argued that an intact ToM mechanism may be essential for developing persecutory delusions, at least in 'pure' delusional disorders, because so-called 'ToM' delusions may only emerge if the capacity to make inferences about the intentions of others is preserved. Similarly, Abu-Akel and Bailey (2000) have suggested that some schizophrenic patients with positive and disorganised symptoms may even have a 'hyper-ToM' by over-attributing knowledge to their interlocutors which they cannot have. These authors propose a continuity model of mentalising abilities in neuropsychiatric disorders, ranging from impaired mentalising, normal mentalising without the ability to apply, and hyper-mentalising with over-attributed mental states or over-generated hypotheses about mental states (Abu-Akel and Bailey, 2000).

These heterogeneous theoretical constructs of mentalising therefore pose some difficulties for a comprehensive review of ToM research in schizophrenia. In addition, compared to the emotion recognition tasks, the methodology of ToM studies is even more diverse. The tests developed to assess ToM capture such divergent aspects as making inferences about intentions derived from hints of indirect speech, understanding metaphor, irony and *faux pas*, first- and second-order false belief, and tactical deception. Moreover, the problem of 'real-life' presentation of the tasks has not satisfactorily been resolved (Simpson, Done and Vallée-Tourangeau 1998). So far, short text passages, partially enacted with simple props, and ToM cartoons involving sorting and completion tasks have been applied, similar to the ToM tests that have been used in autism research (overview in Baron-Cohen 1995).[3] Thus, divergent results may arise, simply because mentalising may be facilitated by judging situations for which the outcome is clear compared to situations in which the outcome has to be anticipated on the basis of inferred mental states of the actors involved. Basically, however, the questions arising with respect to ToM deficits in schizophrenia may be addressed in a similar manner to those in the previous section.

[3] In the 1920s and 1930s, Karl Kleist, a renowned German neuropsychiatrist, conducted similar experiments using picture stories to assess the 'social self' in brain-damaged veterans of World War I (Kleist 1934).

ToM IN SCHIZOPHRENIA — A DIFFERENTIAL DEFICIT OR GENERAL COGNITIVE IMPAIRMENT?

Early studies of ToM in schizophrenia have been somewhat inconclusive regarding the specificity of the predicted ToM deficits. Corcoran, Mercer and Frith's first study (1995) comprised a set of 10 short passages presenting an interaction between two characters. The task was to comprehend the real intention behind indirect speech. The stories were read aloud to the subjects and repeated once on request. The schizophrenic patients were clustered according to Frith's (1992) model. As predicted, the schizophrenic group as a whole scored significantly lower than the controls on the ToM tasks, with the worst performance by subjects with negative and disorganised symptoms. Although there was no correlation between IQ and performance on the ToM tasks in healthy controls, a substantial correlation in the schizophrenic group implies that the latter used generalised cognitive abilities to solve the hinting tasks, which may in turn be interpreted as indirect indicator for a specific mentalising deficit in schizophrenia. More recently, a study by Mitchley *et al.* (1998) similarly addressed the comprehension of irony in schizophrenia. Irony is supposed to require intact mentalising abilities in both the speaker and the listener. The difficulty with irony and metaphor is to extract the intended meaning from speech utterances that often indicate the opposite of the literal meaning. In this study, the irony task comprised brief written scenarios to which one out of three answers had to be attributed. Only one of the answers of choice contained the correct ironical interpretation. The irony tasks were compared to utterances that had to be interpreted literally. It turned out that schizophrenic subjects were impaired in understanding irony relative to psychiatric controls, and that they were more likely to interpret the ironical stories literally. The failure to understand irony was associated with lower IQ and with negative (but not positive) symptoms in the schizophrenic group. However, the difference between the schizophrenic subjects and controls remained significant when IQ was co-varied out. It may therefore be concluded that compromised irony comprehension was not simply a result of general intellectual impairments, but rather a specific deficit related to ToM (Mitchley *et al.* 1998). Frith and Corcoran (1996) had previously examined ToM story comprehension. The text passages were read aloud to the subjects, while cartoon drawings illustrating the passages were presented. The subjects had to answer three first-order and three second-order false-belief questions, as well as reality questions to make sure that the stories were correctly remembered. Patients with behavioural symptoms and paranoid symptoms performed more poorly on ToM tasks than non-psychotic controls. When including only those patients who answered the reality questions correctly, the groups with behavioural symptoms and paranoid symptoms passed the first-order tasks in less than 80% of instances, and paranoid patients passed the second-order

tasks in roughly 60%. However, the drop-out rate in the group with behavioural symptoms was too high for further statistical assessment of second-order false belief tasks, indicating that the second-order ToM stories outreached the general information-processing capacity of the schizophrenic patients (Frith and Corcoran 1996). Consequentially, Corcoran, Cahill and Frith (1997) introduced cartoon drawings into ToM testing, because it had become obvious that the test devices have to be simple, short, and must at best not interfere with other cognitive domains. In this study, the appreciation of visual jokes requiring mental state attribution was assessed. Seven jokes contained a false belief, and three jokes a deception, whereas another set of 10 jokes could be understood simply in 'non-mental' terms. The schizophrenic patients understood the mental state jokes less well than the 'physical' jokes, whereas there was no difference in the control group. Within the schizophrenic group, patients with behavioural symptoms (positive and negative) performed worst on both the physical and the mental state jokes. Patients with paranoid symptoms and, contrary to expectations, patients with passivity symptoms also performed more poorly on the mental state jokes than the controls. However, there was no significant difference in appreciation of physical jokes between these groups. The results remained significant when IQ differences were co-varied out. Patients in remission did not differ from controls in either set of jokes. In conclusion, these three studies partially supported the predictions based on Frith's (1992) model.

Using a modified study design comprising first- and second-order false-belief tasks enacted with props, including ratings of appropriate use of mental state terms in explaining the depicted stories and 'non-mental' tasks matched for complexity, Pickup and Frith (2001) have recently confirmed the findings that IQ, severity of psychopathology and duration of illness do not fully account for the divergent performance on ToM tasks between the symptom subgroups. Rather, a regression analysis revealed that the severity of positive and negative behavioural symptomatology predicted impaired ToM comprehension in this study, suggesting a specific deficit at least in the behavioural symptoms group. Only in paranoid schizophrenics were the difficulties in understanding ToM tasks associated with lower IQ. The authors concluded that paranoid patients have subtle ToM impairments and that those patients with higher IQ could possibly compensate these deficits by using general problem-solving capacities (Pickup and Frith 2001).

Doody *et al.* (1998) assessed ToM story comprehension of schizophrenic patients using the 'Sally-Anne Task' and the 'Ice Cream Van Test', both tasks originally developed for ToM assessment in children (Wimmer and Perner 1983; Perner and Wimmer 1985). Illustrative maps and dolls were additionally presented to facilitate task comprehension. The performance of schizophrenic subjects was compared with that of patients with affective disorders, learning abilities, a sub-group of schizophrenic patients with a co-morbid learning

disability, and with healthy controls. The schizophrenic subjects performed more poorly on the second-order ToM task than the healthy controls, whereas no such difference emerged in performance on the first-order false-belief task. Performance on ToM tasks was associated with positive, negative and general symptomatology only in the schizophrenic group, not in the other groups, therefore lending support for a the assumption of a specific deficit. However, since the schizophrenic patients with a co-morbid learning disability performed more poorly than the 'pure' schizophrenics, a cumulative effect of lowered IQ and psychopathology on ToM abilities could not be ruled out (Doody *et al.* 1998). Likewise, in a similar setting Mazza *et al.* (2001) found differences in ToM performance between schizophrenic patients with reality distortion, psychomotor poverty and disorganisation, respectively, according to Liddle's (1987) three-dimensional model. Schizophrenics with psychomotor retardation performed significantly poorer on first-order false-belief tasks than the other groups. The performance on the second-order false-belief tasks was, however, different, with the disorganised group performing more poorly on one task that required a higher memory load. The differential deficit was not a function of IQ in this study. In sum, the authors concluded that ToM impairments may be specifically linked to negative symptoms in schizophrenia (Mazza *et al.* 2001).

With respect to Hardy-Baylé's proposal (1994), the study group of Sarfati and his co-workers investigated ToM in schizophrenia, putting emphasis on patients with thought and speech disorganisation, compared to non-disorganised schizophrenics, patients with affective disorders and normal controls. Sarfati and colleagues developed ToM tests comprising a series of comic strips, depicting an action in which a volitional state of a character had to be inferred. The subjects were asked to choose one of four answer cards to complete the picture sequence. Only one of the answer cards provided an appropriate ending in light of the character's mental state. A series of studies of ToM (Sarfati *et al.* 1997a,b, 1999) revealed that schizophrenic patients with predominant thought and speech disorganisation performed significantly more poorly on pictorial and verbal ToM tasks than non-disorganised schizophrenic patients and patients with affective disorders, thus pointing to a specific ToM deficit in this subgroup. In one study (Sarfati *et al.* 1997a), the schizophrenic subjects chose answers by chance in some tasks, suggesting that they were impaired in referring to context-specific information. Interestingly, when introduced to a forced-choice paradigm, it turned out that the disorganised schizophrenics no longer chose their answers at random. Instead, their decisions appeared to rely on familiar and unambiguous situations, which the authors interpreted as a compensatory cognitive strategy for ToM deficits (Sarfati and Hardy-Baylé 1999). Patients with a disorganisation syndrome performed worst on false-belief tasks and on intention tasks compared to non-disorganised schizophrenic subjects and controls, but these differences were more prominent in the false-belief tasks. Conventional subtyping of

schizophrenia, i.e. paranoid, disorganised, undifferentiated and residual types, did not reveal any significant differences in task performance in two separate studies, which in part contradicts the findings based on Frith's (1992) model (Sarfati *et al.* 1997a,b).

In a series of experiments in which the pictorial answer cards were replaced by verbal material, the schizophrenic patients with and without disorganisation symptoms improved, as did the affective disorder and normal control groups. However, the differences between the disorganised and the other groups remained significant (Sarfati *et al.* 1999). Positive and negative symptoms, as well as neuroleptic medication and IQ, did not account for these differences. When differentiating between good and poor performers among the schizophrenic subjects, the schizophrenics who improved in ToM performance after introduction of verbal material were, contrary to expectations, not specifically those with the most prominent thought and speech disorganisation. The subgroup who did not benefit from verbalisation, however, had a significantly longer duration of illness compared to the patients with remediated ToM performance and good performers (Sarfati *et al.* 2000). These differences between the groups appeared not to be related to the severity of general psychopathology. In this study, however, the impact of IQ remained ambiguous, since the good performers had a significantly higher IQ compared to the remediable and the poor performers.

To date, the most sophisticated series of experiments to distinguish selective ToM deficits in schizophrenia from general cognitive impairments were conducted by Langdon and colleagues (1997, 2001). In the first study, a complex test battery comprising picture sequencing tasks of 'mechanical', social-script, pretence, unrealised goal, intention and false-belief stories was given to a group of schizophrenic patients and to controls. Patients as a whole performed more poorly than controls on all tasks. Among the schizophrenic patients, however, three subgroups emerged, one of good sequencers, a second group of patients who only made errors on false-belief tasks, and a third with general difficulties in sequencing. These findings point to a selective ToM deficit in the second group. In a subsequent test to differentiate a representational deficit from difficulties in inferring inner states in general, patients were asked to explain the stories. The storytelling was rated according to different types or levels of mental state attribution, i.e. perception, desire, emotion, and intentional and non-intentional cognition. In support of the assumption of a selective ToM deficit in the second group, these patients used terms related to intentional cognition less often. Contrary to predictions, however, the patients with a generalised impairment used emotional and intentional terms less frequently, and instead more often relied on expressions related to perception. Langdon *et al.* (1997) subsequently assessed whether mentalising deficits in schizophrenia were related to impaired self-awareness, by constructing tasks involving recall of pretence, unrealised intentions and guesses. Schizophrenic

patients of the second group performed more poorly than controls on recalling past intentions, supporting the selective mentalising deficit hypothesis. The first group, however, also had some difficulties in recalling past intentions; whereas the third group showed an overall impairment in symbolic representation. When dividing the schizophrenic sample according to Liddle's (1987) categories, patients who performed well on the sequencing tasks (group 1) and patients who made selective errors in false-belief tasks (group 2) did not differ regarding their levels of reality distortion, but the third group scored significantly higher on reality distortion and negative symptoms than the other groups, whereas the second group scored higher on negative symptom ratings than the first group of good sequencers. Moreover, the duration of illness correlated with poor metarepresentational abilities, suggesting a link between negative symptoms and poor mentalising. No correlation, however, was found between reality distortion, e.g. paranoid symptoms, and poor metarepresentation (Langdon *et al*. 1997). In a subsequent study, Langdon *et al*. (2001) focused on the differentiation between selective disruption of a mentalising module from alternative accounts for a ToM impairment in schizophrenia, such as executive planning and disengagement deficits. In addition to the mechanical, social script and ToM sequencing tasks, 'capture' stories depicting misleading cues were given to the subjects, in order to test their abilities to inhibit salient information in favour of less salient details. As predicted, the patients performed more poorly on executive planning tasks, and were significantly impaired in the disengagement and mentalising tasks. In a logistic regression model, however, the patients' performance on false-belief sequencing tasks remained a significant predictor of their patient status, even when executive functions and impairments of inhibitory control were taken into account, therefore underscoring the assumption of a selective disruption of mentalising abilities in schizophrenia (Langdon *et al*. 2001). Contrary to Frith's (1992) model, poor mentalising did not predict any measures of reality distortion, such as paranoid ideation, although higher scores of negative symptoms were best predicted by general picture sequencing deficits and by poor disengagement abilities, which confirmed the findings of the previous study.

An explorative study directly comparing ToM capacities in childhood schizophrenia and autism with normally developing children (Pilowsky *et al*. 2000) has confirmed that mentalising deficits exist in both developmental disorders, but probably to different extents. In a series of tests of understanding beliefs, deception and false beliefs using simulation of 'real-life' situations, schizophrenic children performed more poorly than controls on false-belief tasks, but not on the other tasks. Autistic children, by contrast, were impaired in understanding belief and false belief relative to normally developing children. The main difference between children with autism and those with schizophrenia was the better comprehension of deception by the schizophrenic

subjects, who even outperformed normally developing children. No differences emerged in task performance between paranoid, and disorganised or undifferentiated schizophrenics. It remained unclear, however, whether these differences in ToM abilities were specific to the respective disorder, or whether the same underlying cognitive processes are involved (Pilowsky *et al.* 2000).

In my own explorative study, a sequencing task, a first-order and a second-order false-belief task and a deception task, combined in a single picture story, was given to a sample of patients with chronic disorganised schizophrenia and to a healthy control group (Brüne 2003). The schizophrenic subjects performed more poorly on the sequencing task and on the second-order false-belief task, but not on the first-order false-belief task and the deception task. Also, the total score of the tasks was significantly lower compared to controls. No difference emerged in sequencing a 'physical' story. In line with previous studies (e.g. Langdon *et al.* 1997), the performance on the ToM story was correlated with the duration of the illness. Unexpectedly, however, when controlling for verbal IQ, the difference between schizophrenic subjects and controls disappeared, although there was no difference in psychopathology ratings between the entire schizophrenic group and the IQ-matched sample. Contrary to previous studies, the lack of difference here implies that mentalising abilities may be less impaired in disorganised schizophrenic patients if the test design is modelled in such a way that the outcome of an interaction between two characters can be monitored 'on-line'. Moreover, since the paradigm was assessed using a single set, ceiling effects may have partially accounted for the statistically insignificant differences.[4] In addition, the physical story was not clearly matched for task complexity, hence the question of whether the ToM deficit in the schizophrenic group was specific could not be conclusively resolved (Brüne 2003).

ToM DEFICIT — STATE- OR TRAIT-DEPENDENT?

Most of the above-mentioned studies have demonstrated that poor mentalising abilities in schizophrenia are related to the acuity and symptomatology of the disorder. Patients in remission and patients who solely exhibit passivity symptoms according to Frith's (1992) model have, with few exceptions, performed as well on ToM tasks as controls, findings which support the assumption that a ToM deficit in schizophrenia represents a state rather than a trait variable (e.g. Corcoran, Mercer and Frith 1995; Pickup and Frith 2001). This suggestion has been buttressed by a study addressing whether ToM

[4] Examples of recently developed ToM picture stories comprising understanding of cooperation, first- and second-order false beliefs and tactical deception, based on the explorative data, are shown in Figures 13.1–3 and can be obtained from the author on request, including scoring sheet.

abilities differ during acute episodes of psychosis and periods of recovery (Drury, Robinson and Birchwood 1998). The test battery involved multiple second-order false-belief tasks, the substitution of a co-referential term that tested the ability to represent the mental state of a character in a linguistic context, a metaphor sentence completion task, and the interpretation of metaphor and irony. The acute schizophrenics performed more poorly than non-schizophrenic patients on second-order false-belief tasks and on metaphor tasks, with the performance on the latter being only significantly different when presented in a story but not when part of a sentence completion task. No differences remained after recovery from the acute episode. Moreover, there was no difference between patients with persecutory delusions during the acute phase compared to non-deluded subjects, except for memory, suggesting that ToM deficits are not related to persecutory delusions *per se*. In conclusion, these findings support the notion of state dependence of ToM deficits in schizophrenia. One possible explanation could be that the cognitive plasticity may deteriorate with the chronicity of the schizophrenic disorder (Drury, Robinson and Birchwood 1998).

In contrast, Langdon and Coltheart (1999) have proposed that ToM deficits may rather be trait-dependent vulnerability markers of some psychotic symptoms. To clarify this issue, a group of non-clinical subjects with high schizotypal scores were compared with low schizotypal subjects regarding their mentalising abilities. It had been predicted that, if poor mentalising is a major cause for psychotic symptoms in schizophrenia, and if a continuum exists between schizotypy and schizophrenia, non-clinical individuals with high schizotypal ratings would make more errors than low schizotypics on ToM tests. Moreover, if there is a differential deficit in mental state attribution, schizotypal subjects would perform normally on 'mechanical' and 'social script' sequencing tasks. The study design was similar to the assessment of schizophrenic patients, as outlined above (Langdon *et al*. 1997). In addition, a self-rating exercise of schizotypal traits was given to the participants. As predicted, the high-schizotypics performed more poorly on sequencing false-belief stories compared to low-schizotypics. If the sample was divided into good and poor mentalisers, however, poor mentalisers showed a non-significant trend to score higher on cognitive-perceptual traits, such as magical thinking and unusual perceptual experiences analogous to delusional and hallucinogenic psychotic symptoms. As in the previous study of schizophrenic patients, neither disengagement nor executive planning impairments accounted for the mentalising deficit in high-schizotypal subjects, since they performed as well on the respective tasks as low-schizotypal individuals. The results, therefore, strongly supported the assumption of selective ToM impairment in schizotypal subjects and backed up a continuity model of psychosis, thus supporting the view of a trait-like deficit. Moreover, there was some suggestive evidence that mentalising deficits may play a causal role for the development of

psychotic symptoms, and that such deficits may not solely be explained by social alienation (Langdon and Coltheart 1999).

ToM DEFICIT AND SOCIAL COMPETENCE

The association of social competence with ToM capacities in schizophrenia are far less well examined than the relationship of social skills to emotion recognition in schizophrenia. Some studies, however, give us indirect hints about what should be expected from future work on this topic. Cutting and Murphy (1990), for instance, assessed schizophrenic patients' ability to appreciate social knowledge about their culture, using a Social Knowledge Questionnaire. The subjects were asked to choose one of four answers to questions about how they would react in certain circumstances, which in fact tested their view of the social world. The schizophrenic subjects gave more odd answers to questions with a marked social component than manic and depressed controls, independent of attention deficits, whereas no such difference emerged regarding questions relatively free of social content. The authors concluded that the schizophrenics' social naïveté may be an intrinsic constituent of the illness, rather than being the result of social isolation (Cutting and Murphy 1990).

Similarly, Corcoran and Frith (1996) addressed the conversational conduct of schizophrenic patients and its relationship to psychotic symptomatology. A sample of schizophrenic subjects was compared to a clinical and a non-clinical control group with respect to their ability to apply basic conversational rules and pragmatics, as categorised according to the maxims put forward by Grice (1975), such as quantity, quality, relevance, politeness and tact. Tact and politeness have been argued to be the result of social learning, and closely related to the ability to predict the mental state of a person who receives an impolite or a polite response. It was predicted that schizophrenic patients with negative symptoms would fail to recognise these conversational rules, and that, by contrast, paranoid schizophrenics would be better at applying these rules, with the exception that they would be impaired in behaving politely in novel situations. Overall, the predictions were confirmed: patients with negative symptoms performed worst on all rules in comparison with paranoid, remitted schizophrenic patients and controls, indicating that patients with negative symptoms were ignorant of basic conversational rules. The only exception occurred with the relevance category, where no group differences emerged. Patients with paranoid symptoms only differed from controls with respect to politeness and tact, suggesting a specific deficit in mentalising 'on the spot' in this group (Corcoran and Frith 1996). An interesting case study conducted by Abu-Akel (1999) compared the linguistic skills of two patients suffering from disorganised schizophrenia with healthy persons. He hypothesised that patients, in contrast to healthy subjects, would violate pragmatic and

cohesion rules of conversation, due to a compromised ToM. In this study, cohesion analysis was restricted to the use of reference (phoricity) and the assessment of the use of pragmatics, focusing on the maxims of relation and quantity. The two schizophrenic subjects involved in the study were free of medication. Two loosely structured interviews were conducted with each patient by a clinician who was unaware of the study goals. The interviews were transcribed and analysed according to the variables described above. The analysis revealed that both patients violated the maxims of relation and quantity when compared with healthy control persons. Moreover, with respect to cohesion, it turned out that the schizophrenic patients used unclear references more often than the control subjects. These findings point to the fact that schizophrenics are impaired in monitoring accurately what information is needed by their interlocutors. Since the patients nevertheless tried to cooperate with their interview partners, Abu-Akel interpreted the use of certain cohesive links (so-called 'bridging endophoric references') by the patients during conversation to indicate that the patients referred to some shared reality and that they even assumed that their interlocutors shared their knowledge, suggesting that the patients over-attributed intentions and dispositions to their interlocutors, i.e. they had a 'hyper'-ToM. On the other hand, the incongruent speech of disorganised schizophrenic patients may in part be explained by a lack of frontal inhibition, leading to difficulties in selecting among competing hypotheses of others' mental states (Abu-Akel 1999). The novel aspect of this study is that the appropriate use of language, and hence successful social communication, is ultimately linked to having a ToM.

Another indirect clue comes from a study of interpersonal Machiavellianism in schizophrenia (Sullivan and Allen 1999). The authors assessed the performance of schizophrenic men and women on the Mach-IV scale (Christie and Geis 1970), comprising items dealing with views of human nature, deceitful tactics and morality. Many studies had previously shown that non-clinical men generally score higher on the Mach-IV scale than women. This study revealed a differential picture regarding subjects with schizophrenia. Schizophrenic men scored significantly lower than controls, supporting the notion that schizophrenic men are more socially naïve and thus unconditionally value honesty and morality, which, from an evolutionary perspective, would reduce their success in social competition. Women with schizophrenia, by contrast, scored higher on some items and lower on others. This finding may be interpreted to indicate that schizophrenic women, too, unconditionally value honesty and morality, but at the same time are more suspicious than men. Their total score, therefore, did not differ from controls. Although the schizophrenic sample was not differentiated by subtype, these results may underscore that patients with schizophrenia have reduced social skills related to ToM impairments and application of social rules and tactics (Sullivan and Allen 1999).

CONCLUDING REMARKS AND FUTURE DIRECTIONS

The present chapter has dealt with a group of psychotic disorders traditionally labelled 'schizophrenia', with special emphasis on their symptomatology in terms of social cognition and social behaviour. The term 'social cognition' has been adopted from Leslie Brothers' (1990) definition as:

> 'the processing of any information which culminates in the accurate perception of the dispositions and intentions of other individuals'.

One hallmark of schizophrenic disorders is undoubtedly the compromised social behaviour of the affected individuals, as confirmed in several ethological studies (e.g. Troisi 1999). Moreover, social cognitive models have proved to be equally important, or even superior to, non-social cognitive models, due to their independent contribution of the idea of variance in order to explain schizophrenic symptoms and behaviour (Penn *et al.* 1997). Emotion recognition and ToM undoubtedly represent evolved psychological mechanisms that are crucial for any human social interaction. Many studies have revealed more evidence in favour of than against the assumption that these capacities are specifically impaired in schizophrenia. This does not rule out, however, that non-social cognitive abilities influence the task performance in these domains. On the other hand, it has been criticised that, for example, some of the control tasks designed to differentiate emotion recognition deficits from more global face recognition impairments, such as the Benton Test of Facial Recognition (Benton *et al.* 1978), resemble the emotion recognition tests too closely, and may therefore control for the affective rather than the social quality of stimuli (Penn *et al.* 1997).

Both emotion recognition and ToM likely deteriorate in the course of the illness, although longitudinal studies of first-episode schizophrenic patients are lacking. An alternative account could be that schizophrenic disorders with predominantly negative symptoms reflect primarily underdeveloped social cognitive capacities, similar to autistic disorders (e.g. Pickup and Frith 2001). If so, these patients would be expected to express more childhood precursor symptoms than non-negative schizophrenic patients (e.g. Crow, Done and Sacker 1995). Schizophrenic people with a normally developed ToM, however, may as adults also have more difficulties in mentalising 'on the spot', i.e. determining how and when to apply ToM abilities in novel situations (Corcoran and Frith 1996; Brüne 2003). This assumption is also supported by the finding that people with schizophrenia probably make less use of strategic social thinking than do controls (Sullivan and Allen 1999). This interesting issue needs further examination, because clinical impressions also suggest that patients with schizophrenia very rarely deceive intentionally in therapeutic settings, in contrast, for instance, to patients with personality disorders.

The majority of studies suggests that emotion recognition and ToM abilities in schizophrenia represent state rather than trait variables, although other researchers have found the opposite. This issue requires therefore further evaluation (Langdon and Coltheart 1999; Edwards *et al.* 2001). Likewise, emotion recognition and ToM deficits in schizophrenia have a profound influence on social functioning. The results concerning this topic, however, are still somewhat inconclusive, and this matter, as far as ToM is concerned, has only been assessed indirectly (e.g. Corcoran and Frith 1996; Poole, Tobias and Vinogradov 2000).

Furthermore, the relationship of emotion recognition to ToM in schizophrenia has not been directly addressed so far. This is critical, however, because these two abilities to perceive and to interpret social signals in terms of intentions and dispositions of other individuals are closely inter-related and have certainly co-evolved in primates and human beings. A recent study of ToM and emotion recognition abilities in autistic children, children with pervasive developmental disorder, children with attention deficit/hyperactivity disorder, and other psychiatric and non-psychiatric controls has shown non-differential impairments in these domains across groups, with autistic children performing worst (Buitelaar *et al.* 1999). Thus, one would predict similar unidirectional deficits in schizophrenic disorders. Additional evidence comes from functional imaging and primate studies of the underlying brain systems involved in emotion recognition and mentalising. The data suggest the existence of a neural network connecting the medial prefrontal cortex, the anterior cingulate and the superior temporal sulcus (the 'dorsal' system), and linking the orbitofrontal cortex and regions next to the amygdala (the 'ventral' system) (Frith and Frith 1999, 2001). The former of these brain areas are most important for mentalising and are also involved in self-monitoring and perception of biological motion (but not of movements of inanimate objects), whereas the latter are crucial for emotion recognition and recognition of other individuals. The neural system responsible for mentalising possibly evolved, therefore, from the ability to predict the actions of other individuals, particularly of conspecifics (Castelli *et al.* 2000).

This model, however, implies that a dissociation or functional disruption of the processes of emotion recognition and mentalising is theoretically conceivable, such that the function of the phylogenetically older system, i.e. emotion recognition, may be preserved, while the 'younger' ToM mechanism is impaired, and vice versa. In most 'natural' interactions, however, emotion recognition may facilitate mental state attribution, but may not be obligatory for it, whereas the reverse does not necessarily apply.

Although speculative to date, this hypothesis is consistent with many clinical symptoms of schizophrenia, e.g. a disintegration of emotional display and the content of the subjectively felt emotion (parathymia) may arise from disparate functioning of emotion recognition and self-representation. Likewise, impaired

irony and proverb (metaphor) comprehension and 'concretistic' thinking in schizophrenia may result from such a functional disruption of emotion recognition and mentalising. This occurs because, for the correct understanding of irony and metaphor, reading between the lines requires not only mental state attribution but also correct interpretation of the tone of voice and of the deliberately mismatched or suppressed facial expression of the ironic actor. Moreover, as emotions decisively influence cognitive processes, e.g. via so-called 'somatic markers', by recollecting previous emotional experiences (Damasio 1996; Charlton, in this volume), empathy may be indispensable for mental perspective taking in 'real-life' situations (Brothers 1989). It would therefore be interesting to investigate whether there are disparate emotion recognition and mentalising deficits in schizophrenic patients with parathymic affect, and whether patients with prominent affective flattening differ from those with positive symptoms, such as delusions and hallucinations.

We have to concede from the review of studies of emotion recognition and mentalising in schizophrenia that their comparability is limited. One probable reason is that the diverse experimental designs probably cover different aspects of emotion recognition and of ToM abilities. Findings from both studies of acute stages and chronic schizophrenia suggest that deficits of these capacities emerge following a hierarchical model of breakdown, such that the ontogenetic order of acquisition of differentiation is reversed. This assumption is supported by the fact that social learning becomes increasingly important for decoding subtle social signals and for advanced 'mind-reading', which in turn suggests that socially learned abilities are more vulnerable to dysfunction than more 'innate' capacities. With respect to emotion recognition, this would imply that context-dependent ambiguous facial expressions of emotions, are most difficult to interpret, followed by complex expressions of emotions, such as surprise, disgust, shame and contempt. By contrast, fear, anger, happiness and sadness are probably more easy to understand, although a difference between positive and negative emotion comprehension may exist (Bell, Bryson and Lysaker 1997; Buitelaar *et al.* 1999; Blair *et al.* 1999). Similarly, ToM impairments would, in the first place, manifest themselves through misapprehension of *faux pas* and metaphor, deception and understanding of false belief, in the reverse order of complexity.[5] In some psychotic disorders, however, the basic ToM functions may be preserved, although misinterpretation of the intentions of others may be compromised with respect to attributional styles (see Chapter 15). This would be consistent with the conclusion of Walston *et al.* (2000), that an intact ToM mechanism is essential for developing persecutory

[5] Dunbar has suggested that sixth-order intentionality may be the limit of understanding written ToM levels, e.g. V thinks that W believes that X has the impression that Y imagines that Z intends... (Dunbar 1998).

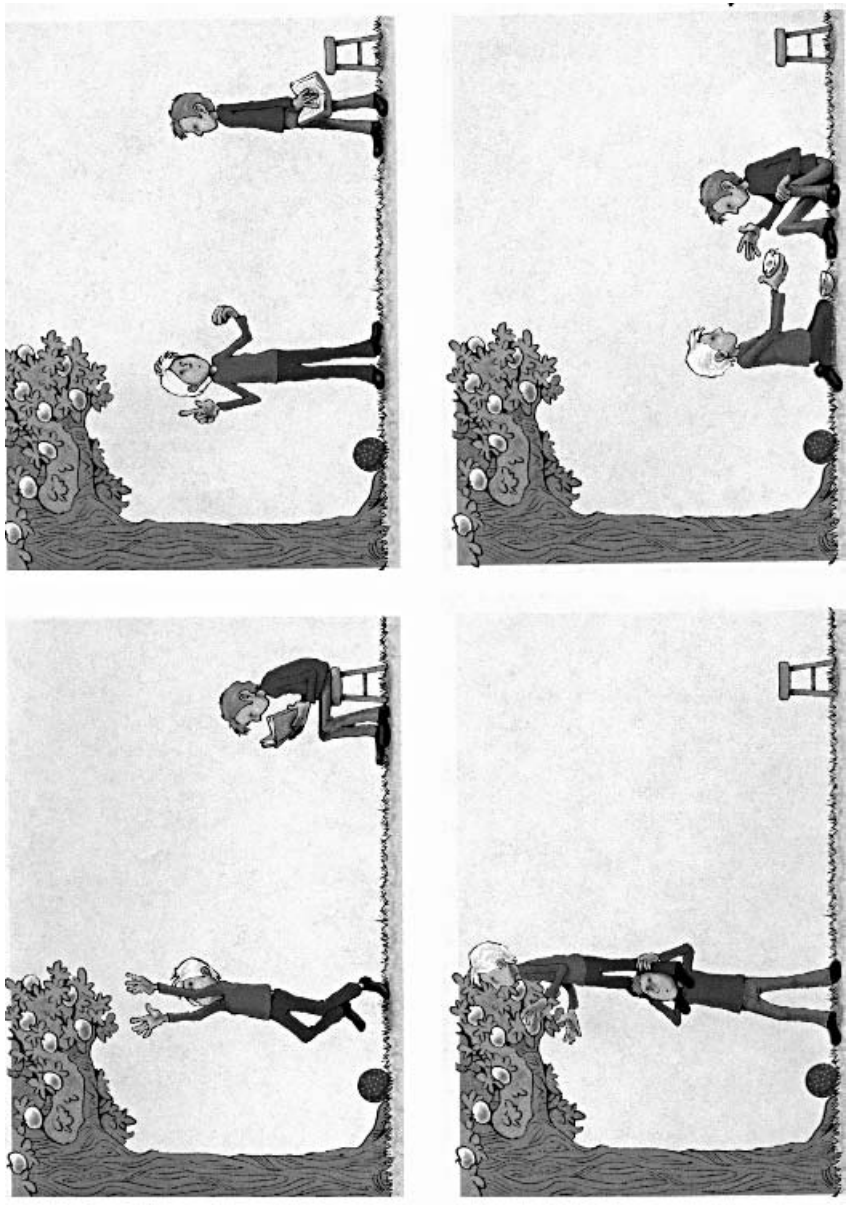


Figure 13.1. Cartoon test series: cooperative action on the basis of mutual agreement

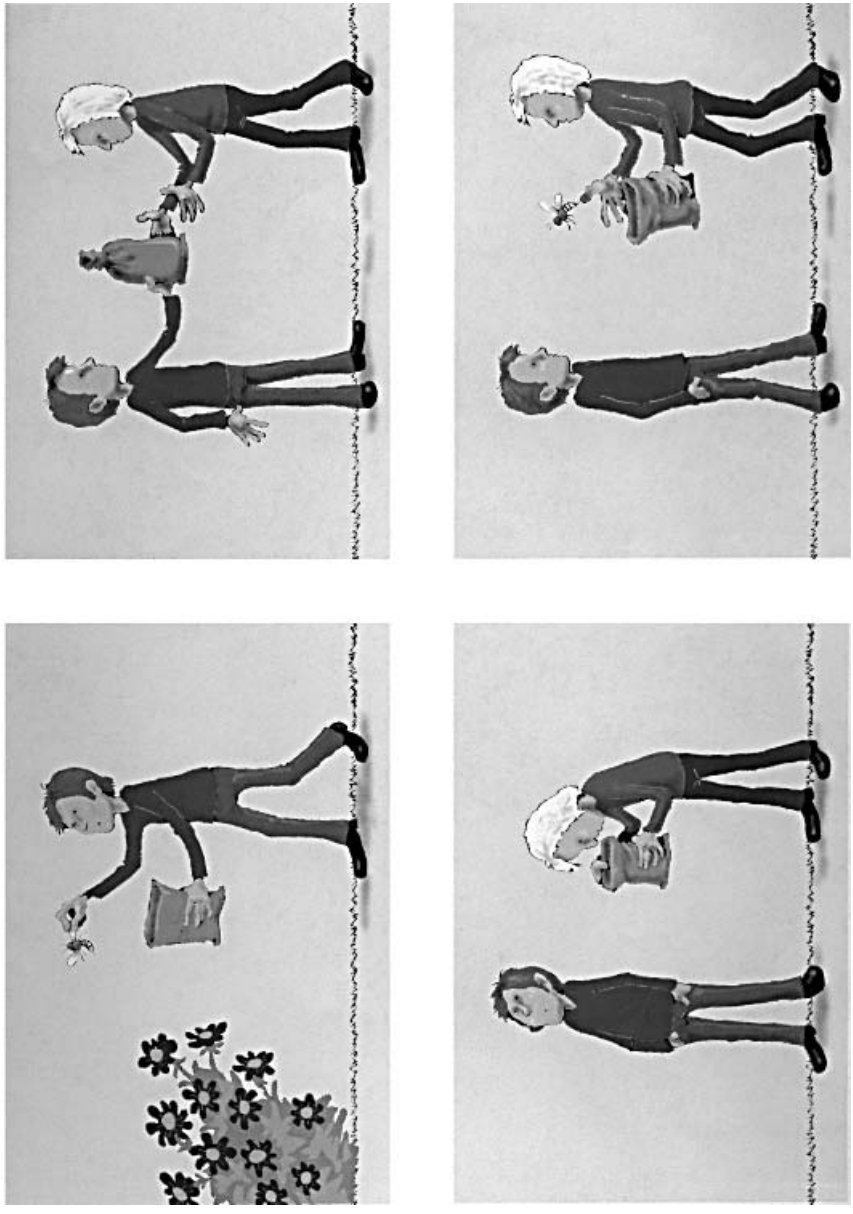


Figure 13.2. Cartoon test series: intentional deception

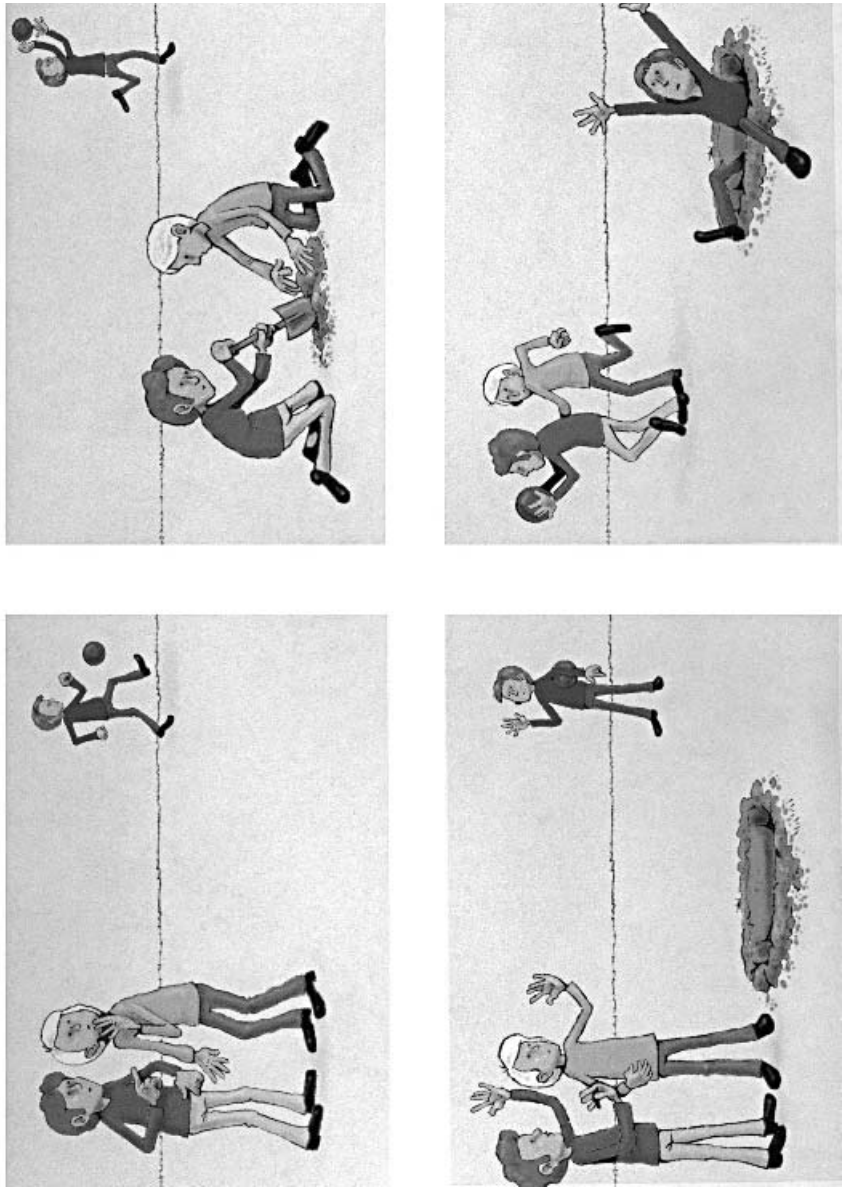


Figure 13.3. Cartoon test series: cooperation to deceive

delusions, and that some patients even over-attribute intentions to others (Abu-Akel and Bailey 2000).

From a therapeutic perspective — provided that emotion recognition and ToM are related to social competence — it is crucial to evaluate whether patients could benefit from cognitive training in these domains (Penn *et al.* 1997; Sarfati 2000).

APPENDIX

Figures 13.1–13.3 show examples of a cartoon series, each cartoon consisting of four pictures. The ToM tests involve a sequencing task and understanding of belief, intention, false belief and deception, according to different levels of intentionality. Figure 13.1 depicts a situation in which two characters act cooperatively on the basis of mutual agreement; in Figure 13.2, one character intentionally deceives a second person; and in Figure 13.3, two persons cooperate in order to deceive a third character. Ratings of the standardised ToM questions are made on separate scoring sheets.

REFERENCES

Abu-Akel, A. (1999). Impaired theory of mind in schizophrenia. *Pragm. Cogn.* **7**, 247–282.

Abu-Akel, A. and Bailey, A.L. (2000). The possibility of different forms of theory of mind. *Psychol. Med.* **30**, 735–738.

APA (American Psychiatric Association) (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn. American Psychiatric Association, Washington, DC.

Baron-Cohen, S. (1995). *Mindblindness: an Essay on Autism and Theory of Mind.* Bradford MIT Press, Cambridge, MA.

Bell, M., Bryson, G. and Lysaker, P. (1997). Positive and negative affect recognition in schizophrenia: a comparison with substance abuse and normal control subjects. *Psychiat. Res.* **73**, 73–82.

Bellack, A.S., Morrison, R.L., Wixted, J.T. and Mueser, K.T. (1990). An analysis of social competence in schizophrenia. *Br. J. Psychiat.* **156**, 809–818.

Bellack, A.S., Blanchard, J.J. and Mueser, K.T. (1996). Cue availability and affect perception in schizophrenia. *Schizophr. Bull.* **22**, 535–544.

Benton, A., van Allen, M., Hamser, K. and Levin, H. (1978). *Test of Facial Recognition Manual.* Benton Laboratory of Neuropsychology, Iowa City, IO.

Blair, R.J.R., Morris, J.S., Frith, C.D., Perrett, D.I. and Dolan, R.J. (1999). Dissociable neural responses to facial expressions of sadness and anger. *Brain* **122**, 883–893.

Bleuler, E. (1911). *Dementia praecox oder die Gruppe der Schizophrenien.* Deutike, Leipzig.

Brothers, L. (1989). A biological perspective on empathy. *Am. J. Psychiat.* **146**, 10–19.

Brothers, L. (1990). The social brain: a project for integrating primate behavior and neurophysiology in a new domain. *Concepts Neurosci.* **1**, 27–51.

Brüne, M. (2001). Social cognition and psychopathology in an evolutionary perspective — current status and proposals for research. *Psychopathology* **34**, 85–94.
Brüne, M. (2002). Towards an integration of interpersonal and biological processes: evolutionary psychiatry as an empirically testable framework for psychiatric research. *Psychiatry* **65**, 48–57.
Brüne, M. (2003). Theory of mind and the role of IQ in chronic disorganized schizophrenia. *Schizophr. Res.* **60**, 57–64.
Bryson, G., Bell, M. and Lysaker, P. (1997). Affect recognition in schizophrenia: a function of global impairment or a specific cognitive deficit. *Psychiat. Res.* **71**, 105–113.
Bryson, G., Bell, M., Kaplan, E., Greig, T. and Lysaker, P. (1998). Affect recognition in deficit syndromes in schizophrenia. *Psychiat. Res.* **77**, 113–120.
Buitelaar, J.K., van der Wees, M., Swaab-Barneveld, H. and van der Gaag, R. (1999). Theory of mind and emotion recognition functioning in autistic spectrum disorders and in psychiatric control and normal children. *Dev. Psychopathol.* **11**, 39–58.
Castelli, F., Happé, F., Frith, U. and Frith, C. (2000). Movement and mind: a functoinal imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage* **12**, 314–325.
Christie, R. and Geis, F.L. (eds) (1970). *Studies in Machiavellianism*. Academic Press, New York.
Conrad, K. (1979). *Die beginnende Schizophrenie. Versuch einer Gestaltanalyse des Wahns*, 5th edn (first published 1958). Thieme, Stuttgart.
Corcoran, R., Mercer, G. and Frith, C.D. (1995). Schizophrenia, symptomatology and social inference: investigating 'theory of mind' in people with schizophrenia. *Schizophr. Res.* **17**, 5–13.
Corcoran, R. and Frith, C.D. (1996). Conversational conduct and the symptoms of schizophrenia. *Cogn. Neuropsychol.* **1**, 305–318.
Corcoran, R., Cahill, C. and Frith, C.D. (1997). The appreciation of visual jokes in people with schizophrenia. A study of 'mentalizing' ability. *Schizophr. Res.* **24**, 319–327.
Cosmides, L. and Tooby, J. (1994). Beyond intuition and instinct blindness: toward an evolutionarily rigorous cognitive science. *Cognition* **50**, 41–77.
Cramer, P., Weegmann, M. and O'Neil, M. (1989). Schizophrenia and the perception of emotions. How accurately do schizophrenics judge the emotional states of others? *Br. J. Psychiat.* **155**, 225–228.
Crow, T.J. (1985). The two-syndrome concept: origins and current states. *Psychol. Bull.* **11**, 471–486.
Crow, T.J., Done, D.J. and Sacker A. (1995). Childhood precursors of psychosis as clues to its evolutionary origins. *Eur. Arch. Psychiat. Clin. Neurosci.* **245**, 61–69.
Cutting, J. (1981). Judgement of emotional expression in schizophrenics. *Br. J. Psychiat.* **139**, 1–6.
Cutting, J. and Murphy, D. (1990). Impaired ability of schizophrenics, relative to manics or depressives, to appreciate social knowledge about their culture. *Br. J. Psychiat.* **157**, 355–358.
Damasio, A.R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Phil. Trans. R. Soc. Lond. Ser. B* **351**, 1413–1420.
Darwin, C. (1872). *The Expression of Emotions in Man and Animals*. Murray, London.
Davies, M. and Stone, T. (1995). *Mental Simulations: Evaluations and Applications*. Blackwell, Oxford.
Doody, G.A., Götz, M., Johnstone, E.C., Frith, C.D. and Owens, D.G. (1998). Theory of mind and psychoses. *Psychol. Med.* **28**, 397–405.

Dougherty, F.E., Bartlett, F.S. and Izard, C.E. (1974). Response of schizophrenics to expressions of the fundamental emotions. *J. Clin. Psychol.* **30**, 243–246.

Drury, V.M., Robinson, E.J. and Birchwood, M. (1998). 'Theory of mind' skills during an acute episode of psychosis and following recovery. *Psychol. Med.* **28**, 1101–1112.

Dunbar, R.I.M. (1998). The social brain hypothesis. *Evol. Anthropol.* **6**, 178–190.

Edwards, J., Pattison, P.E., Jackson, H.J. and Wales, R.J. (2001). Facial affect and affective prosody recognition in first-episode schizophrenia. *Schizophr. Res.* **48**, 235–253.

Ekman, P. and Friesen, W.V. (1976). *Pictures of Facial Affect*. Consulting Psychologists Press, Palo Alto, CA.

Federman, E.J., Drebing, C.E, Zaref, J.I. and Oepen, G. (1998). Processing in recognition of facial affect in schizophrenia. *Percept. Motor Skills* **87**, 484–486.

Feinberg, T.E., Rifkin, A., Schaffer, C. and Walker, E. (1986). Facial discrimination and emotional recognition in schizophrenia and affective disorders. *Arch. Gen. Psychiat.* **43**, 276–279.

Friesen, W.V. and Ekman, P. (1984). EMFACS-7: Emotional Facial Action Coding System, Version 7. Unpublished.

Frith, C.D. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Erlbaum, Hove.

Frith, C.D. and Corcoran, R. (1996). Exploring 'theory of mind' in people with schizophrenia. *Psychol. Med.* **26**, 521–530.

Frith, C.D. and Frith, U. (1999). Interacting minds — a biological basis. *Science* **286**, 1692–1695.

Frith, U. and Frith, C.D. (2001). The biological basis of social interaction. *Curr. Dir. Psychol. Sci.* **10**, 151–155.

Gaebel, W. and Wölwer, W. (1992). Facial expression and emotional face recognition in schizophrenia and depression. *Eur. Arch. Psychiat. Clin. Neurosci.* **242**, 46–52.

Grant, E.C. (1968). An ethological description of non-verbal behaviour during interviews. *Br J. Med. Psychol.* **41**, 177–183.

Grice, H.P. (1975). Logic and conversation. In P. Cole and J.L. Morgan (eds), *Syntax and Semantics*, vol. 3, *Speech Acts*. Academic Press, New York.

Hardy-Baylé, M.C. (1994). Organisation de l'action, phénomènes de conscience et représentation mentale de l'action chez des schizophrènes. *Actual. Psychiat.* **20**, 393–400.

Höschel, K. and Irle, E. (2001). Emotional priming of facial affect identification in schizophrenia. *Schizophr. Bull.* **27**, 317–327.

Ihnen, G.H., Penn, D.L., Corrigan, P.W. and Martin, J. (1998). Social perception and social skill in schizophrenia. *Psychiat. Res.* **80**, 275–286.

Izard, C. (1959). Paranoid schizophrenic and normal subjects' perception of photographs of human faces. *J. Consul. Psychol.* **23**, 119–124.

Izard, C. (1971). *The Face of Emotion*. Appleton-Century-Crofts, New York.

Jones, I.H. and Pansa, M. (1979). Some non-verbal aspects of depression and schizophrenia occurring during the interview. *J. Nerv. Ment. Dis.* **167**, 402–409.

Kerr, S.L. and Neale, J.M. (1993). Emotion perception in schizophrenia: specific deficit of further evidence of generalised poor performance? *J. Abnorm. Psychol.* **102**, 312–318.

Kleist, K. (1934). *Gehirnpathologie*. Johann Ambrosius Barth, Leipzig.

Kohler, C.G., Bilker, W., Hagendoorn, M., Gur, R.E. and Gur, R.C. (2000). Emotion recognition deficit in schizophrenia: association with symptomatology and cognition. *Biol. Psychiat.* **48**, 127–136.

Kraepelin, E. (1899). *Psychiatrie. Ein Lehrbuch für Studirende und Aerzte*, 6th edition. Johann Ambrosius Barth, Leipzig.
Krause, R., Steimer, E., Sänger-Alt, C. and Wagner, G. (1989). Facial expression of schizophrenic patients and their interaction partners. *Psychiatry* **52**, 1–12.
Langdon, R., Michie, P.T., Ward, P.B., McConaghy, N. *et al*. (1997). Defective self and/or other mentalising in schizophenia: a cognitive neuropsychological approach. *Cogn. Neuropsychiat*. **2**, 167–193.
Langdon, R. and Coltheart, M. (1999). Mentalising, schizotypy, and schizophrenia. *Cognition* **71**, 43–71.
Langdon, R., Coltheart, M., Ward, P.B. and Catts, S.V. (2001). Mentalising, executive planning and disengagement in schizophrenia. *Cogn. Neuropsychiat*. **6**, 81–108.
Lewis, S.F. and Garver, D.L. (1995). Treatment and diagnostic subtype in facial affect recognition in schizophrenia. *J. Psychiat. Res*. **29**, 5–11.
Liddle, P.F. (1987). Schizophrenic symptoms: cognitive performance and neurological dysfunction. *Psychol. Med*. **17**, 49–57.
Mazza, M., DeRisio, A., Surian, L., Roncone, R. and Casacchia, M. (2001). Selective impairments of theory of mind in people with schizophrenia. *Schizophr. Res*. **47**, 299–308.
McGuire, M.T. and Polsky, R.H. (1979). Behavioral changes in hospitalized acute schizophrenics. An ethological perspective. *J. Nerv. Ment. Dis*. **167**, 651–657.
McGuire, M.T. and Polsky, R.H. (1983). Sociospatial behavioral relationships among hospitalized psychiatric patients. *Psychiat. Res*. **8**, 225–236.
Mitchley, N.J., Barber, J., Gray, Y.M., Brooks, N and Livingston, M.G. (1998). Comprehension of irony in schizophrenia. *Cogn. Neuropsychiat*. **3**, 127–138.
Mueser, K.T., Doonan, R., Penn, D.L., Blanchard, J.J. *et al*. (1996). Emotion recognition and social competence in chronic schizophrenia. *J Abnorm. Psychol*. **105**, 271–275.
Mueser, K.T., Penn, D.L., Blanchard, J.J. and Bellack, A.S. (1997). Affect recognition in schizophrenia: a synthesis of findings across three studies. *Psychiatry* **60**, 301–308.
Novic, J., Luchins, D.J. and Perline, R. (1984). Facial affect recognition in schizophrenia: is there a differential deficit? *Br. J. Psychiat*. **144**, 533–537.
Penn, D.L., Mueser, K.T., Spaulding, W., Hope, D.A. and Reed, D. (1995). Information processing and social competence in chronic schizophrenia. *Schizophr. Bull*. **21**, 269–281.
Penn, D.L., Spaulding, W., Reed, D. and Sullivan, M. (1996). The relationship of social cognition to ward behavior in chronic schizophrenia. *Schizophr. Res*. **20**, 114–132.
Penn, D.L., Corrigan, P.W., Bentall, R.P., Racenstein, J.M. and Newman, L. (1997). Social cognition in schizophrenia. *Psychol. Bull*. **121**, 114–132.
Perner, J. (1991). *Understanding the Representational Mind*. MIT Press, Cambridge, MA.
Perner, J. and Wimmer, H. (1985). 'John thinks that Mary thinks . . .'. Attribution of second-order beliefs by 5–10 year-old children. *J. Exp. Child Psychol*. **39**, 437–471.
Pickup, G.J. and Frith, C.D. (2001). Theory of mind impairments in schizophrenia: symptomatology, severity and specificity. *Psychol. Med*. **31**, 207–220.
Pilowsky, T., Yirmiya, N., Arbelle, S. and Mozes, T. (2000). Theory of mind abilities of children with schizophrenia, children with autism, and normally developing children. *Schizophr. Res*. **42**, 145–155.
Pitman, R.K., Kolb, B., Orr, S.P. and Singh, M.M. (1987). Ethological study of facial behavior in non-paranoid and paranoid schizophrenic patients. *Am. J. Psychiat*. **144**, 99–102.

Ploog, D. (1999). Evolutionsbiologie der Emotionen. In H. Helmchen, F. Henn, H. Lauter and N. Sartorius (eds), *Psychiatrie der Gegenwart*, 4th edn, vol. I. Springer, Berlin, pp. 525–553.

Poole, J.H., Tobias, F.C. and Vinogradov, S. (2000). The functional relevance of affect recognition errors in schizophrenia. *J. Int. Neuropsychol. Soc.* **6**, 649–658.

Rosenthal, R., Hall, J.A., DiMatteo, M.R., Rogers, P.L. and Archer, D. (1979). *Sensitivity to Nonverbal Communication: the PONS Test*. Johns Hopkins University Press, Baltimore, MD.

Salem, J.E., Kring, A.M. and Kerr, S.L. (1996). More evidence for generalized poor performance in facial emotion perception in schizophrenia. *J. Abnorm. Psychol.* **105**, 480–483.

Sarfati, Y. (2000). Deficit of the theory-of-mind in schizophrenia: clinical concept and review of experimental arguments. *Can. J. Psychiat.* **45**, 363–368.

Sarfati, Y., Hardy-Baylé, M.C., Nadel, J., Chevalier, J.F. and Widlöcher, D. (1997a). Attribution of mental states to others in schizophrenic patients. *Cogn. Neuropsychiat.* **2**, 1–17.

Sarfati, Y., Hardy-Baylé, M.C., Besche, C. and Widlöcher, D. (1997b). Attribution of intentions to others in people with schizophrenia: a non-verbal exploration with comic strip. *Schizophr. Res.* **25**, 199–209.

Sarfati, Y., Hardy-Baylé, M.C., Brunet, E. and Widlöcher, D. (1999). Investigating theory of mind in schizophrenia: influence of verbalization in disorganized and non-disorganized patients. *Schizophr. Res.* **37**, 183–190.

Sarfati, Y. and Hardy-Baylé, M.C. (1999). How do people with schizophrenia explain the behaviour of others? A study of theory of mind and its relationship to thought and speech disorganisation in schizophrenia. *Psychol. Med.* **29**, 613–620.

Sarfati, Y., Passerieux, C. and Hardy-Baylé, M. (2000). Can verbalization remedy the theory of mind deficit in schizophrenia? *Psychopathology* **33**, 246–251.

Schneider, F., Gur, R.C., Gur, R.E. and Shtasel, D.L. (1995). Emotional processing in schizophrenia: neural behavioural probes in relation to psychopathology. *Schizophr. Res.* **17**, 67–75.

Scholl, B.J. and Leslie, A. (1999). Modularity, development and 'theory of mind'. *Mind Lang.* **14**, 131–153.

Shaw, R.J., Dong, M., Lim, K.O., Faustman, W.O. *et al.* (1999). The relationship between affect expression and affect recognition in schizophrenia. *Schizophr. Res.* **37**, 245–250.

Shimizu, T., Shimizu, A., Yamashita, K., Iwase, M., Kajimoto, O. and Kawasaki, T. (2000). Comparison of eye-movement patterns in schizophrenic in normal adults during examination of facial affect displays. *Percept. Motor Skills* **91**, 1045–1056.

Simpson, J., Done, J. and Vallée-Tourangeau, F. (1998). An unreasoned approach: a critique of research on reasoning and delusions. *Cogn. Neuropsychiat.* **3**, 1–20.

Spitzer, R.L., Endicott, J. and Robins, E. (1975). *Research Diagnostic Criteria*. New York State Psychiatric Institute, New York.

Staehelin, B. (1953). Gesetzmäßigkeiten im Gemeinschaftsleben schwer Geisteskranker. *Schweiz. Archiv Neurol. Psychiat.* **72**, 277–298.

Streit, M., Wölwer, W. and Gaebel, W. (1997). Facial affect recognition and visual scanning behaviour in the cause of schizophrenia. *Schizophr. Res.* **24**, 311–317.

Suddendorf, T. and Whiten, A. (2001). Mental evolution and development: evidence for secondary representation in children, great apes, and other animals. *Psychol. Bull.* **127**, 629–650.

Sullivan, R.J. and Allen, J.S. (1999). Social deficits associated with schizophrenia in terms of interpersonal Machiavellianism. *Acta Psychiat. Scand.* **99**, 148–154.

Sweet, L.H., Primeau, M., Fichtner, C.G. and Lutz, G. (1998). Dissociation of affect recognition and mood state from blunting in patients with schizophrenia. *Psychiat. Res.* **81**, 301–308.

Toomey, R., Schuldberg, D., Corrigan, P. and Green, M.F. (2002). Nonverbal social perception and symptomatology in schizophrenia. *Schizophr. Res.* **53**, 83–91.

Troisi, A., Spalletta, G. and Pasini, A. (1998). Non-verbal behaviour deficits in schizophrenia: an ethological study of drug-free patients. *Acta Psychiat. Scand.* **97**, 109–115.

Troisi, A. (1999). Ethological research in clinical psychiatry: the study of non-verbal behavior during interviews. *Neurosci. Biobehav. Rev.* **23**, 905–913.

Walston, F., Blennerhassett, R.C. and Charlton, B.G. (2000). 'Theory of mind', persecutory delusions and the somatic marker mechanism. *Cogn. Neuropsychiat.* **5**, 161–174.

Whiten, A. (2000). Social complexity and social intelligence. *Novartis Found. Sympos.* **233**, 185–196.

Whittaker, J.F., Deakin, J.F.W. and Tomenson, B. (2001). Face processing in schizophrenia: defining the deficit. *Psychol. Med.*, **31**, 499–507.

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

Wölwer, W., Streit, M., Polzer, U. and Gaebel, W. (1996). Facial affect recognition in the course of schizophrenia. *Eur. Arch. Psychiat. Clin. Neurosci.* **246**, 165–170.

14

Theory of Mind Delusions and Bizarre Delusions in an Evolutionary Perspective: Psychiatry and the Social Brain

BRUCE G. CHARLTON
Department of Psychology, University of Newcastle upon Tyne, UK

THE SOCIAL BRAIN AND PSYCHIATRY

The phrase 'social brain' embodies the idea that the problems of living in a complex social group have been a dominant selection pressure in recent human evolutionary history (Humphrey 1976; Byrne and Whiten 1988; Brothers 1990). One consequence is that many distinctively human behaviours can be linked with adaptations for social living.

The perspective of 'the social brain' has particular relevance to psychiatry, since 'psychiatric symptoms (e.g. hallucinations, delusions, phobias, obsessions) are frequently dominated by social content, and a disruption of social relationships is highly characteristic of psychiatric illness. Indeed, it might plausibly be argued that the distinctive nature of many psychiatric illnesses — that thing that makes them 'psychiatric' — may be the combination of emotional pathology with social impairment. Certainly, the emotional and the social are intimately related at the level of brain function, since social reasoning depends upon evolved brain systems for monitoring and modelling emotional responses to social scenarios (Damasio 1994; Charlton 2000).

The following chapter will demonstrate how a human social evolution has been used to clarify and refine the diagnostic category of delusions. I will argue

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

that two distinct types of delusions may be discriminated: 'theory of mind' delusions and 'bizarre' delusions.

DELUSIONAL DISORDER

Delusional disorder (DD) may be described as a psychiatric condition in which a delusion is the primary symptom and patients are otherwise 'normal' (Charlton and McClelland 1999). The subject matter of the delusion is variable; with persecutory, jealous, grandiose, erotomaniac and somatic subtypes being recognised (APA 1994). Furthermore, delusional disorder is characterised by specifically social abnormalities of behaviour, such as morbid states of jealousy, love, self-awareness or fear of other people (Charlton and McClelland 1999).

Delusions are false beliefs. However, this is an insufficient definition, and further attempts at definition are all somewhat unsatisfactory (Garety and Hemsley 1994). Usually, a delusion is defined as a false belief that is also strongly held, such that it exerts a strong influence on behaviour and is not susceptible to counter-arguments or counter-evidence (or, at least, the delusion is unshakeable over a short timescale and in the absence of systematic attempts at belief modification). Furthermore, in order to distinguish delusions from 'religious' beliefs, a delusion is also supposed to be out of context with the usual cultural beliefs for that society (Sims 1995).

Delusional disorder is an unusual diagnosis in general psychiatric practice, and delusions are most frequently seen to occur as only one element in more complex clinical syndromes. In other words, most delusions are observed along with other 'psychotic' symptoms, such as hallucinations or incoherent speech ('thought disorder') as part of one of the classic syndromes of 'madness' such as schizophrenia, mania, psychotic depression and 'organic' symptoms indicative of generalised brain dysfunction, such as dementia or delirium.

But in delusional disorder false beliefs occur largely in isolation as 'encapsulated' delusions. Such individuals do not have other primary psychological symptoms, such as hallucinations, incoherent speech or qualitatively abnormal mood states (although the delusions may lead to secondary symptoms, e.g. a belief in persecution may lead to secondary emotional change, such as fear or anger specifically in relation to the imagined persecutors). It therefore seems likely that the majority of people diagnosable with delusional disorder are never seen by psychiatrists (Charlton and McClelland 1999; Walston, Blennerhassett and Charlton 2000).

THEORY OF MIND (ToM)

It has recently been noted that the subject matter of delusional disorders is distinctive, since the false beliefs are 'social' in content, and typically concerned with the assumed dispositions, motivations and intentions (DMIs) of other

people. For this reason, delusional disorders have been labelled 'theory of mind' (ToM) delusions, because they seem to involve inferences (or theories) about what is going on in the minds of other people (Charlton and McClelland 1999). For example, the commonest types of delusional disorder are probably those relating to jealousy over sexual infidelity and those in which there is a false belief of persecution. Both are considered to be commoner in men. Jealous delusions may involve a man believing that his wife is concealing from him that she is having an affair with another man, while persecutory delusions typically involve a man believing that he is the victim of a hostile plot to attack and probably kill him. It has been suggested that these delusions derive from errors of inference relating to the contents of other people's minds.

Since delusional disorder is related to reasoning about the dispositions, motivations and intentions of other people, the mechanism by which such reasoning is performed in humans requires consideration. ToM is the ability, displayed by adult humans, to make inferences about the content of other people's minds. Beyond this bald statement, conceptualisations of ToM vary widely between published accounts, and in different branches of biology. I suggest (for reasons argued elsewhere; Charlton 2000) that the essence of ToM, its central adaptive importance and the reason why it evolved, is that the ToM mechanism is primarily concerned with making inferences concerning the dispositions, motivations and intentions of other people. It is not, therefore, a mechanism that evolved for making theories about the 'factual' or 'knowledge' content of other people's minds, although in language-using humans the ToM mechanism may be used for this purpose.

The ToM mechanism enables other people's behaviour to be interpreted in the light of inferred DMIs. This is necessary because many human social behaviours are ambiguous unless interpreted with knowledge of 'intent', e.g. a clenched fist can be a threat, a salute or a gesture of encouragement, according to the motivation of the fist-wielder. Discrimination between these different meanings requires an understanding of the social context of behaviour, including the individual differences between human beings and the different ways in which these different human beings interact. For example, only by knowing the DMIs of others can we distinguish between friends and foes, know who to trust and who to avoid — and in general build those alliances that underpin human society — especially in those small-scale tribal societies in which humans evolved (Walston, David and Charlton 1998).

STRATEGIC SOCIAL INTELLIGENCE

The ability to make inferences concerning the DMIs of others can be termed *strategic* social intelligence (Charlton and McClelland 1999; Charlton 2000). It is strategic because it is used in planning future social strategies. Strategic social

intelligence is therefore a kind of 'internal modelling', a way in which the brain can run 'simulations' of possible future scenarios and evaluate their outcomes. Most animals do not have strategic social intelligence used for planning, although many animals have highly developed *tactical* social intelligence (detecting and responding to social cues) for dealing with the here-and-now of face-to-face interactions within the same species.

Strategic social intelligence (SSI) is an adaptation which is unique to humans and (probably) a few other species of recently evolved mammals with relatively large brains and complex social systems. SSI is almost certainly found in chimpanzees, bonobos and other great apes, probably in dolphins and elephants, and perhaps in others. SSI is valuable in humans (and other species with broadly similar social systems) exactly because here-and-now behaviour cannot be interpreted at face value. No matter how expert we are at reading facial expressions, gestures and vocal intonation, we are still vulnerable to being deceived if we do not take into account the context of the social interaction and any evidence of the DMIs of the individuals concerned. For example, a smiling, charming and plausible stranger may knock on your door and offer to give you €1000 tomorrow if you will give him €100 today. It is strategic social intelligence that enables you to infer that this man cannot necessarily be taken at face value, and that he is more likely to be a confidence trickster than a benefactor. Damasio (1994) has shown that people with some types of neurological impairment, such as pre-frontal cortex damage and non-dominant parietal cortex damage, lack of just this kind of social understanding. They demonstrate serious impairments of planning and judgement, especially in relation to social interactions, and hence would be vulnerable to exploitation by deception, e.g. they would probably take the above-described confidence trickster at face value.

STRATEGIC SOCIAL INTELLIGENCE AND THEORY OF MIND

Strategic social intelligence is — like most biological systems designed by natural selection — useful but imperfect, and the ToM mechanism for inferring the content of other people's minds is not 100% reliable. Humans do not have direct access to the content of other people's minds, and predicting the outcomes of social interactions is inevitably probabilistic. The validity of inferences about other minds can only be checked against the subsequent behaviour of individuals by (implicitly) asking the question: is subsequent behaviour consistent with the assumptions concerning dispositions, intentions and motivation?

Yet even this check on the validity of ToM inferences is flawed, since inferences about DMIs affect our interpretation of behaviour. For instance, if we make an inference that a person is aggressively hostile towards us, then we may look at her subsequent behaviour to discover whether this inference is correct. Yet the suspicion of hostility affects our interpretation of this

behaviour. If we have already decided that Big Boris is hostile, our emotional state as we observe him is likely to be affected by this inference, and if subsequently he raises a fist towards us we are more likely to be frightened already and to interpret the fist as a threat. We may take the clenched fist as confirmation of Big Boris's aggressive hostility. Yet the interpretation of Boris's gesture as confirmation of a hostile intent may be a mistake if our original 'theory' of his mental state was incorrect.

The above example illustrates how humans make inferences about other people's minds on the basis of information in their own minds. Some of this information involves propositional 'knowledge' about the external world perceived through the senses, but some of the information involves monitoring the *internal* world of emotions, i.e. monitoring body states; both are necessary for understanding other humans.

THE SOMATIC MARKER MECHANISM

Humans use the state of their own bodies in understanding the contents of other minds (Damasio 1994). From the work of Damasio and colleagues, it emerges that the somatic marker mechanism (SMM) is the primary mechanism for making ToM inferences ('somatic' refers to body, and 'marker' to the linking of body state information with perceptions, as described below). Hence, the SMM underpins the function of strategic social intelligence.

The SMM was originally described following studies of defective 'social intelligence' in neurological patients, especially those with pre-frontal syndromes and non-dominant parietal lobe damage. The essence of the SMM is that we make inferences about the DMIs of other people by monitoring our own emotions. So we interpret another person's DMIs in the light of the emotions that this person induces in us. Patients with neurological damage that renders them unable to experience or monitor emotion will lose strategic social intelligence. They may be able to respond appropriately to here-and-now (tactical) social situations, but they cannot understand social context, predict social outcomes or plan adaptive social interactions. They have lost the ability to perform internal modelling of human interactions, to run 'simulations' and evaluate probable outcomes.

Emotions are cognitive representations of body states (remembering that 'body' states also include brain states; Damasio 1999). So the emotion of fear is the state of a body that has been activated in a specific fashion by the sympathetic nervous system, with raised heart rate, blood flow diverted to muscles, erected hair, etc. The brain continuously monitors these body states, and the emotion of fear comprises the brain representations of this body state. The SMM uses information on emotions in order to evaluate the significance of imagined social scenarios.

AN ILLUSTRATION OF THE SMM IN ACTION

As an illustration, imagine that you perceive a large, aggressive man called Big Boris who is approaching you. You respond with the emotion of fear, which means that your sympathetic nervous system prepares your body for action, and the brain monitors this activated state and takes the fear into account when generating appropriate behaviour in response to the situation.

Your perception of the identity of Big Boris is integrated with the body state of fear in response to his approach, and this combined perception and emotion is laid down in long-term memory. So, the memory now contains a combined perceptual–emotional representation of 'Big Boris-fear'. The body state is the 'somatic' part (*soma* = body) and it is used as a 'marker' for the perception, hence the name 'somatic marker mechanism'. 'Fear' is the marker for 'Big Boris', and the distinctive character of the marker is that when the memory of Big Boris is reactivated, this also reactivates the emotional marker.

This type of long-term memory becomes important in SSI, e.g. when planning future interactions in relation to Big Boris. Whenever you simulate or internally model interactions with Big Boris, you will recall his identity. And in recalling Boris's identity you will also re-enact the associated emotional 'marker', in this instance, the body state of fear that is linked to the identity of Boris. This re-enacted fear will influence your decisions in relation to Big Boris, so that you are more likely to consider him as an enemy to be avoided or eliminated than as a potential friend or ally. Your inferences concerning Big Boris's dispositions, motivations and intentions will therefore be based upon your emotional response to him.

This description of the SMM demonstrates that the basis of ToM is the ability to use our own emotional states to make theories about the DMIs of other people. No doubt there are other mechanisms involved in human ToM inferences, especially since humans have abstract symbolic language, but Damasio's work demonstrates that the SMM is the fundamental basis of the ToM ability — without the SMM, strategic social intelligence is severely impaired. ToM is not, therefore, based on knowledge of other people's thoughts, but on assumptions about their nature and purposes. In other words, recalling Big Boris is fear-inducing for you, and your inference is that, because Boris is a fear-inducing man, he is probably hostile.

A CASE STUDY OF PERSECUTORY DELUSIONS

This model of the ToM mechanism was developed partly as a consequence of studying individuals who suffered that type of delusional disorder characterised by persecutory delusions. A case study of persecutory delusional disorder was undertaken in order to establish the characteristics of such individuals and to examine their reasoning processes (Walston, Blennerhassett and Charlton 2000).

'Pure' cases of persecutory delusions (i.e. patients validly diagnosable as having delusional disorder) are not easy to find, since such people only come to the attention of psychiatrists when the patients becomes psychiatrically ill due to the consequent fear of harm, or else when the patient becomes a danger to himself or others due to his reaction to the delusion.

Cases were sought by enquiry concerning the caseloads of 17 psychiatrists and four community psychiatric nurses; 34 cases were referred, but only four fulfilled the study criteria of being 'pure' cases without other potentially obvious confounding psychiatric or neurological symptoms. The four cases were all men, aged 32–43, who had suffered severe persecutory delusions for periods ranging from 6 months to 10 years. All were confirmed (by psychometric tests) to be of normal intelligence, with intact cognitive functioning and without any *other* formal psychiatric diagnosis. It is important to emphasise that the mental functioning, personality, conversation and social demeanour of these men was entirely normal, except in relation to the subject matter of the delusion.

It was found that all four men were able to lead normal social lives apart from their delusions, having good relations with family and friends — in other words, their persecutory delusions were restricted to a specific group of supposed persecutors, and did not encompass everyone in the world. Furthermore, the subjects were all able to perform a range of visual and verbal social reasoning tasks, so-called ToM tests, to a very high standard.

Through tape recorded interviews, a detailed account of their delusional beliefs was elicited and the results were clear-cut. Delusions of persecution were restricted to a specific group of imagined persecutors and to the hostile intentions of this group. All four men believed that they were being persecuted by a violent gang of male strangers. The delusional belief was restricted to the specific social category of people in the gang and the specific fear of persecution. The reasoning processes that led to these delusions were not irrational, even though they were incorrect. It appeared that these subjects were systematically misinterpreting real but ambiguous information in the light of inferred hostile intent. For example, one of the subjects saw a stranger enter his club, and assumed that the stranger was a member of the gang he thought was watching him; or the same subject saw someone carrying a bag and made the assumption that this bag contained a gun. In other words, the stimuli were real, but were interpreted in the light of pre-existing assumptions concerning the hostile intentions of presumed persecuting gang members. It was the initial assumption that these subjects were being pursued by a hostile gang that was incorrect. Aside from this initial assumption, their interpretations of events was rational even when it was highly implausible.

Fear of violent attack from gangs of male strangers was probably consistent with a common source of threat in ancestral human environments (Walston, David and Charlton 1998). Certainly, evidence from twentieth

century hunter-gatherers, as well as chimpanzees, implies that alliances of unrelated males are a significant cause of premature death.

INTERPRETATION OF THE CASE STUDY

Delusional disorders seem to be a consequence of logical reasoning from false premises concerning other people's mental states. They are based on false assumptions rather than logical errors. There may, in principle, be many causes leading to a person making false assumptions concerning other people's mental states, given that the system for making inferences relies upon monitoring subjective emotional states. But whatever the cause in each individual case, once established, such delusional beliefs appear able to sustain themselves by a circularity in reasoning, based on specific emotions being linked to specific social categories, e.g. misinterpretation generates fear, which in turn leads to a misinterpretation, which causes more fear.

Since delusional disorders are aspects of SSI which involve the ToM mechanism, it is possible to explain many of the clinical and phenomenological features of delusional disorder on the basis of understanding the nature of the ToM mechanism. The case study of persecutory delusions seems to confirm that delusional disorders occur in a context of non-pathological cognitive functioning, including an *intact* ToM mechanism.

Indeed, the subject matter of several types of delusional disorders seems specifically to be associated with important social challenges to reproductive success in the probable ancestral environment. Delusions are concerned with such matters as alliances of enemies (persecutory), fidelity of sexual partners (jealous) and vital questions such as other people's perceptions of one's own status (grandiose), appearance (somatic) and sexual attractiveness (erotomania). These represent some of the main categories of social competition in the human ancestral environment, in other words the main evolutionary selection pressures, and therefore some of the main functions of the 'social brain'.

CHARACTERISTICS OF ToM DELUSIONS

SOCIAL SUBJECT MATTER OF DELUSIONAL DISORDERS

Humans are social animals, and the reproductive success of our ancestors depended crucially upon their ability to negotiate the social milieu and compete with members of their own species. Furthermore, human psychological mechanisms evolved under tribal conditions with small-scale, face-to-face social interactions; and presumably they functioned well, on average, in these circumstances. But these same mechanisms now operate in a mass social environment populated mainly by strangers performing frequently unobserved

acts, and in these conditions what was adaptive may become pathological (Charlton 1998, 2000).

The subject matter of delusional disorder bears a striking resemblance to the principal categories of social interaction that have evolutionary importance and require mental state inferences. In other words, delusional disorders apparently reflect the nature of social selection pressures in an ancestral environment, e.g. homicide is a major cause of premature male death and hence failure to reproduce under tribal conditions (Wrangham and Peterson 1996), and many homicides are the result of 'gangs' of males. Persecutory aggression by hostile alliances of unrelated males was probably, therefore, a highly significant feature of ancestral social life, and it makes sense that inferences concerning persecution by male alliances have the potential to act as a powerful influence on behaviour (Walston, Blennerhassett and Charlton 1998).

Similarly, a conjectural evolutionary scenario to account for erotomanic and some somatic delusions can be derived from theories of human sexual selection (Buss 1994; Miller 2000). The major variable that influences a man's attractiveness to women is status, and erotomania can be seen as a condition in which a woman becomes delusionally attracted to an unattainable but high-status male (Enoch and Trethowan 1991; Mullen and Pathe 1994). This question has recently been examined systematically in a study of 246 cases of erotomania (Brüne 2001). Evolutionary predictions drawn from 'Sexual Strategies Theory' were confirmed by the pattern of reported symptoms. By contrast, a woman's physical attractiveness to men is primarily a matter of physical beauty (cues of youth and health; Buss 1994) and in the somatic type of delusional disorder, a common presentation is in a hypersensitive, insecure woman of reproductive age who has become preoccupied that she is physically unattractive due to some bodily impairment (e.g. a foul odour) or personal ugliness (e.g. blemished skin, large nose). Somatic delusions of this type are reported to be unusual in women beyond reproductive age, and when somatic delusions of this type are found in men, it could be predicted that they will be more common among those who rely on their appearance for attracting sexual partners, e.g. homosexual men, or men of lower social status.

Such explanations are obviously highly general, and fail to account for the occurrence of pathologies in specific individuals. Nonetheless, they make predictions about individuals diagnosed as suffering from psychiatric states, and these predictions can be tested, e.g. in studies such as the case study of persecutory delusions summarised above.

FALSE BELIEFS ARE UNAVOIDABLE WHEN INFERRING MENTAL STATES

The false beliefs found in delusional disorder are social, and involve mistaken mental state inferences, i.e. misjudging the dispositions, motivations and

intentions of other people. Such mistakes are inevitable, given the nature of the ToM mechanism. Beliefs concerning the mental state of others cannot always be true because beliefs cannot be checked against objective criteria — there is no direct access to other minds.

And some circumstances might plausibly make such mistakes more probable. Inferences concerning the state and content of other minds depend upon information from one's own subjective emotional responses. In other words, accurate knowledge of another mind depends on knowledge of one's own body state. When a person's subjective emotional response is inappropriate or pathological, then the inference of mental state will probably be wrong. In principle, almost any cause of increased fear might be expected to predispose to persecutory delusions if there was already a somatic marker linked to a specific class of persons. Such fear may be an aspect of personality or circumstance, or due to a disease process, drug side effects or withdrawal effect. For instance, I have described the emergence of persecutory delusions following the withdrawal of a neuroleptic (Charlton 2000).

The relationship between specific emotional states and specific errors in inference leads to testable predictions, although understanding of the typology of emotional states remains incomplete. Persecutory delusions may specifically be associated with particular personality types, including certain attributional styles (Bentall, Kinderman and Kaney 1994). Another possible link between emotions and delusional disorder may also be seen in the reported association between low self-esteem (i.e. perceived low status) and morbid jealousy (Mullen and Martin 1994).

BELIEFS CONCERNING ToM INFERENCES WILL BE RESISTANT TO COUNTER-ARGUMENT

Beliefs concerning the state of mind of other people may be powerfully resistant to counter-argument, *despite* the fact that such beliefs arise from potentially insecure inferences. It might perhaps have been expected that such potentially insecure beliefs based on subjective emotional factors would be only weakly held — but in fact these are exactly the kind of beliefs held with greatest passion.

Presumably this paradox is explained by the fact that the human social domain is intrinsically competitive (Byrne and Whiten 1988; Whiten and Byrne 1997). Indeed, it is suggested that the ToM mechanism evolved as a direct consequence of exactly this human vs. human competition. Deception and concealment of hostile motivations and damaging intentions can be expected in the social domain.

Therefore, mistrustfulness is adaptive when it comes to judging the social explanations and reassurances of other people. Dishonesty from other people, in these matters, is to be anticipated, since most humans are potential rivals for limited resources (especially matings; Buss 1994; Miller 2000), and none of

them share exactly the same social agenda. It therefore makes sense that beliefs concerning ToM will be neither labile nor readily abandoned, despite their subjective basis. In a rivalrous social world where no-one can wholly be trusted, each person must reach his/her own conclusions about the motivations, disposition and relationships of other people, and must rely on his/her own judgements, however imperfect and unreliable.

ToM DELUSIONS ARE ENCAPSULATED DUE TO THE NATURE OF THE ToM MECHANISM

When mental state delusions are a consequence of the ToM mechanism (i.e. dependent on the SMM) they depend upon a cognitive representation that incorporates a social identity with an emotion (Damasio 1994, 1999).

The somatic marker is an emotion linked to a social category. When the memory of that social category (e.g. Big Boris, or a specific gang of drug dealers) is activated, then the body state linked to it is re-enacted. In other words, because an emotion is linked specifically with a category of social identity that elicits that emotion, mental state inferences will be restricted to the particular person or group described by that social category.

This potentially explains why pure cases with delusions of persecution can nevertheless maintain friendly and cooperative social relationships with people outside of the social category of their presumed persecutors — people outside the specific social category do not elicit activation of the somatic marker emotion of fear. This seems to apply to both men and women, although the social categories differ. Probably, female persecutory delusions usually relate to familiar people, while male delusions relate to strangers (Walston, David and Charlton 1998). Analogously, female jealousy is mainly concerned with commitment of love and resources, while male jealousy is mainly concerned with sexual infidelity (see below). In both instances, the delusion is encapsulated, although the general social category differs between the sexes.

A CASE OF MORBID JEALOUSY

Having developed the idea of ToM delusions by concentrating on the evidence provided by persecutory delusions, clinical case histories concerning delusions of sexual infidelity were examined to check whether they conformed to the features described.

'Morbid jealousy syndrome' describes a condition of inappropriate or excessive jealousy, specific to the sexual partner, and which dominates behaviour; this becomes delusional when it involves a false belief in the sexual infidelity of the spouse or sexual partner. Morbid jealousy can occur in a pure form (i.e. without the presence of another psychiatric diagnosis) in both males

and females, although it is commoner in males (Shepherd 1961; Enoch and Trethowan 1991). What follows is a true account, although names and identifying details have been changed (Charlton 2000).

Edward is a man in his mid-20s. He had an uneventful childhood, was an average pupil and left school without taking examinations to serve an apprenticeship. Edward's personality is cautious and careful, and people have commented on his neatness, punctuality and conscientiousness. Although somewhat shy, he has plenty of friends and an active social life. Indeed, he has strong attachments to his family, and a powerfully developed sense of personal responsibility. There is no history of psychiatric illness or current sign of psychiatric illness.

In his early 20s, Edward began a relationship with a younger girl called Frances that lasted several years. As the relationship progressed it became more stormy, with arguments centring around Frances's desire for more freedom to go out with friends, and Edward's increasingly possessive attitude to her and his criticisms of her sexually provocative style of dress. Edward became increasingly worried that Frances might be 'seeing other boys and having sex. If she had sex with anyone else I could never have her back'. The worry escalated into a tormenting preoccupation, and on one occasion Edward was driven to phone one of Frances's friends to check that she was not seeing anyone else; on another occasion he went around the local night clubs to check on her whereabouts.

The situation became so bad that the relationship split up (a 'trial separation'). However, Edward became even more distressed. One evening, Edward saw Frances in a bar, talking to a group of men and dancing in what seemed a provocative fashion. He left the bar ruminating on the possibility that she was seeing other men, and the thought 'jumped through' his mind that she may have had sex with them — although he pushed the thought aside. In an overwrought mood, he waited outside Frances's home in a car to discuss their relationship. She sat by him in the car, an argument broke out and Frances tried to make it up by kissing Edward; but Edward exploded into sudden anger at her sexually provocative manner — and he strangled Frances to death.

Edward was immediately overwhelmed with remorse, drove for miles, and made a determined attempt at suicide. The interview took place in prison, where Edward was awaiting trial for murder.

SEXUAL JEALOUSY IN EVOLUTIONARY CONTEXT

Jealousy in humans is a cultural universal, a complex and characteristic pattern of behaviour in response to specific cues, which serves an adaptive function concerned with paternal investment in offspring. Across the animal kingdom, jealous behaviour is found when males contribute resources to their offspring (especially after birth) and is a response to the problem of uncertain paternity

in species where females potentially mate with more than one male (Wilson and Daly 1992; Buss 1994; Pagel 1997).

Jealousy in men can be seen as an evolved psychological adaptation that operates to reduce the chance of sexual infidelity in a partner, and reduce the chance of misdirected investment (even the act of mating typically requires substantial investment of resources, and loss of the opportunity to invest these in courting other mates; Buss 1994). If a male were to tolerate sexual infidelity and continue to invest resources into a rival male's offspring, he would incur the 'double' genetic penalty of both failing to reproduce and 'wasting' resources on assisting a rival's reproduction. Humans have few offspring, each requiring substantial investment of resources — any child sired by another man represents the loss of a substantial proportion of expected reproductive capacity (Buss 1994).

Jealousy in women is significantly different in its motivation and intentions, since female mammals do not suffer from uncertainty as to the identity of their offspring, and sexual infidelity *per se* is not a problem. The problem for a female is to secure investment to help in rearing offspring, and jealousy is primarily concerned with ensuring that the male partner directs his investment efforts towards the woman's own offspring. So female jealousy is less concerned with the *act* of sexual infidelity and more with the danger of a male partner transferring his affections (and resources) to another female (Wiederman and Allgeier 1993; Buss 1994). Hence, selection pressures have led to different cues that stimulate the emotion of jealousy in men and women: men primarily fear *physical* infidelity (the partner having sexual intercourse with another man) while women primarily fear *emotional* infidelity (the partner falling in love with another woman) (Townsend 1995; Geary *et al*. 1995; Mullen and Martin 1994).

The extreme of *morbid* jealousy, such as displayed by 'Edward' in the case history above, would not usually be considered adaptive (see Sheets, Fredendall and Claypool 1997), since it could severely damage reproductive success, e.g. when it causes the break-up of a relationship or death of one or both partners by homicide (Mowat 1966). However, it remains possible that the threat or possibility of such extreme sanctions may serve as an effective deterrent; hence even intense jealousy may be adaptive on average, or under ancestral conditions.

Jealous delusions be considered as consequences of the ToM mechanism since, although the content may be complex and varied, in pure cases the delusions seem to be consequences of *internally-modelled* social relationships and mental state inferences. Jealous delusions are not about what is happening here-and-now, but instead about what did happen, is happening elsewhere, or might happen in the future. In other words, delusions of jealousy are provoked by *imagined* social interactions, 'simulated' scenarios of sexual infidelity.

Delusions of sexual infidelity are consistent with the four predicted characteristics of ToM delusions. For instance, false-positive (or inappropriate) jealousy is inevitable at a certain frequency, since imaginative construction of possible scenarios cannot always be based upon, or checked against, reality. Also, inferences concerning the intentions of a sexual partner are not directly accessible but can only be checked against behaviours whose interpretation is ambiguous. Jealousy is notoriously resistant to reassurance or counter-argument. There is often no objectively convincing way to contradict the delusional belief. This arises from the fact that jealousy evolved in a context of social competition, where deception is expected as an element of that competition. Nonetheless, false beliefs of sexual infidelity are compatible with being encapsulated and specific to the sexual partner (Enoch and Trethowan 1991). The encapsulation occurs on the basis that the ToM mechanism involves a cognitive linkage between a particular social category and a particular emotion — outside that subject matter and that emotion, cognitive life may proceed relatively unaffected.

BIZARRE DELUSIONS

Not all delusions are ToM delusions. Another important category of false belief is 'bizarre' delusions. Bizarre delusions include many of the most typical delusions seen in classic 'schizophrenic' patients (Sims 1995), e.g. those 'primary' delusions in which a person suddenly becomes convinced of something false without any understandable logical link — 'The traffic lights turned green and I knew I was the son of God', or 'My thoughts stopped and I realised that they were being drawn out of my head by X-rays'.

Some bizarre delusions arise from hallucinations or other abnormal bodily or mental experiences, e.g. 'the voices' may have told a person that he was the son of God, or he may believe that the funny feelings in his abdomen were caused by telepathy. These are bizarre ways of explaining bizarre experiences. However, whether bizarre delusions are primary or secondary, I will argue they share the common cause of global brain dysfunction. Hence, bizarre delusions have very limited relevance to the study of 'the social brain' — emphasising the importance of distinguishing between these two types of delusion.

BRAIN DYSFUNCTION AND DELUSIONS

Everybody has experienced the kind of illogical thinking that leads to bizarre delusions, since this kind of progression of ideas happens in dreams. For instance, dreams may resemble the following: 'I walked into the street and saw a lion and realised that to escape I needed to open a trapdoor hidden underneath the hedge, and the trapdoor opened onto another planet with purple skies and no gravity, but the lion had changed into a flowerpot . . .'

If we awaken from a doze, or just as we are dropping off to sleep, we may recall that our thoughts were 'falling apart' and becoming illogical; this is a brief state of delirium due to 'clouded consciousness'. The release from a normal coherent progression to a quasi-pathological and unpredictable association of ideas varies in severity on a continuum from occasional lapses of concentration to gross incoherence. The process can be observed by other people when a delirious patient exhibits a fluctuating state of consciousness, lucid and rational intervals interspersing drowsy or agitated periods of illogical thought.

Incoherent thinking and illogical reasoning is the consequence of 'clouded' consciousness or generalised brain damage. Whenever thinking is impaired by 'organic' insult to the brain — when drowsy, pyrexial, when the brain is impaired by drugs or alcohol (or by withdrawal from drugs or alcohol), or has extensive pathological damage from dementia, or has been damaged by trauma — then under such circumstances there is a greatly increased potential for impaired reasoning to lead to false beliefs.

Beliefs resulting from illogical thinking can be extremely bizarre, partly because the stream of consciousness is disrupted, and partly because the mechanisms for testing ideas for plausibility and consistency with other ideas are also damaged.

CHARACTERISTICS OF BIZARRE DELUSIONS

Bizarre delusions may be distinguished from ToM delusions in terms of several contrasting psychopathological criteria (Charlton 2000):

1. *Bizarre delusions may have any subject matter*. Whereas ToM delusions are always about social phenomena; bizarre delusions might be about any subject or thing: social or environmental, physical or metaphysical, natural or supernatural.
2. *Bizarre delusional beliefs may survive* objective *refutation*. ToM delusions stem from inferences concerning the mental states of other people; in other words, ToM delusions arise from indirect inferences about entities that are not directly observable. Hence they may not be possible to refute because there is no direct access to other people's mental states, no objective way of demonstrating the dispositions, motivations or delusions of other people. But bizarre delusions may be held *despite* the evidence of direct observation.

   Evidence that any normal person would find compelling is not necessarily persuasive to someone with bizarre delusions. Because the reasoning processes are themselves impaired in bizarre delusions, then a chain of argument that would usually be considered to be conclusive evidence against a belief does not carry the force necessary to compel a change of belief.

   For instance, a person with psychotic depression and nihilistic delusions may believe that his/her internal organs have rotted away, leaving him/her

hollow. Such a person is holding a belief in the existence of a state of affairs that is incompatible with human life — yet this 'fact' of the delusion being impossible is not taken to be compelling. Indeed, this kind of patient may deny that he/she is alive at all, which again contradicts what would be considered to be the possibilities of objective fact.

Whatever arguments or evidence that are brought to bear, the bizarre delusional belief may remain unshaken, because when brain function is impaired we cannot follow logic and chains of reasoning are disrupted, so even objective evidence does not have the power to persuade.

3. *Bizarre delusions are not encapsulated by social category*. ToM delusions are characterised by false beliefs confined to a particular social category, as when the deluded person is only jealous of his wife (but not his sister), or only afraid of the local drug gang (but not the Freemasons). But since bizarre delusions are caused by impaired reasoning, bizarre delusions are not restricted to particular social categories, and delusional thinking is liable to be a feature of many domains of discourse, e.g. bizarre delusions with a persecutory theme may encompass not just a specific group of persecutors, but the whole of humankind in a 'conspiracy' against the subject.
4. *Pure cases of bizarre delusions will not exist*. While ToM delusions can occur as 'pure cases' in people who are otherwise normal, it would be predicted that there will be *no pure cases* of bizarre delusions — that is to say, there will be no cases of people who have an encapsulated bizarre delusion with otherwise normal psychological functioning. In lay terms, all people with bizarre delusions will be overtly 'mad' or in some other way suffering from global brain impairment, inevitably leading to a *variety* of psychological symptoms.

Because bizarre delusions are a consequence of impaired reasoning processes, and impaired reasoning processes will be a consequence of global brain impairment such as delirium or dementia, then when bizarre delusions are observed there will *always* be a widespread impairment in brain function. Such a person will produce not just a single false belief, but a variety of psychological symptoms typical of that form of impairment. A person with bizarre delusions will not merely have a false belief, he/she will also exhibit symptoms such as impairments in concentration, altered mood, and poor performance on short-term memory tasks, consistent with a diagnosis of delirium. Bizarre delusions will therefore *only* be found as part of a psychiatric syndrome, *never* as pure cases.

BIZARRE DELUSIONS CAUSED BY ORGANIC BRAIN IMPAIRMENT

The category of *bizarre* delusions is not explicable in terms of rationally misinterpreting normal perceptions on the basis of misattributed intentions,

motivations, or dispositions. Bizarre delusions require either that the patient is rationally misinterpreting pathological psychological features, such as hallucinations, or else the actual logical processes are irrational due to pathology, e.g. 'They threw an egg at my window, and this meant I was a homosexual, so I switched on my radio'.

It is uncontroversial that irrational thinking and abnormal psychological experiences are a common feature of *organic* brain disease, e.g. delirium, epilepsy or dementia. When a brain is dysfunctional or damaged, then it is unsurprising that the brain cannot perform cognitive processing in the normal fashion. The same applies to sleep or near-sleep states. False beliefs are to be expected in a circumstance when brain impairment has affected the cognitive processes by which beliefs are generated. There is reason to suppose that bizarre delusions are caused by global brain impairment, perhaps most commonly due to chronic and severe sleep disruption.

BIZARRE DELUSIONS, DELIRIUM AND ILLUSORY DREAMS

Chronic severe sleep loss and other forms of sleep disruption are significant clinical features in many psychotic patients, although they are seldom considered as potential aetiological factors for psychotic symptoms. Yet chronic, severe sleep loss can certainly cause delirium, and the probable mechanisms of this link have recently been elucidated (Charlton and Kavanau 2002).

Memory circuits of the brain are reinforced during sleep, a process in which synaptic strengths are maintained at dedicated levels (Kreuger and Obal 1993; Kavanau 1994). Synaptic strength maintenance occurs largely through the action of self-generated, spontaneously occurring, slow brain waves (waves at frequencies less than about 14 cycles/s). This occurs during both rapid-eye-movement (REM) and non-REM sleep, although there are significant differences in function between the two phases. Maintenance is necessary, since all synaptic strengths weaken with time due to 'turnover' of essential molecules. Without remedial action during sleep, all memory circuits that were not being regularly 'exercised' by frequent use while awake would gradually deteriorate and their encoded memories be lost (Kavanau 1996; Stickgold 1998).

Some of the memories being reinforced during sleep rise to the level of 'unconscious' awareness, and these are the memories that provide the substance of our dreams. If our stored memories were valid in every respect, then our dreams would consist solely of 'replays' of past or plausible events in our lives (roughly 85–95% of dream contents derive from authentic contents; Antrobus and Bertini 1992). On the other hand, when memory circuits are faulty, the events and perceptions in the resulting dreams may be distorted or largely illusory. The sources of illusory dream contents are at least three-fold: first, there are disordered synaptic strengths due to normal

imperfections in the processes that store and maintain memories; second, there are the abnormal influences of pathologically altered brain waves; and third, there are the influences of sleep loss leading to incomplete refreshment of synaptic strengths.

The earliest manifestations of sleep deprivation, even after a single night, include significant impairment of cognitive performance and changes in mood. After only 48 h most subjects report illusions and/or visual and tactile hallucinations, and these become more intense as deprivation progresses (Everson 1997). The pathology underlying organic delirium additionally involves abnormalities in slow waves, observable by EEG and characterized by deviations from normal frequency, form, magnitude or distribution (Slaby and Cullen 1987). Such pathological waves are presumably incapable of reinforcing memory circuits in the usual fashion. With a cumulative weakening of synaptic strengths in affected circuits, subsequent recall of such distorted memories produces hallucinations, delusions and other hallmarks of delirium.

These alterations of mental state apparently reflect the use of incompetent circuitry that accumulates during the extended periods of sleep loss and/or the effects of pathological brain waves. Just as illusory dreams in normal individuals often are the result of activation of incompetent circuits, the symptoms of organic delirium probably owe their origin primarily to the activation of incompetent circuitry of functionally pathological origin. And these mechanisms lead to the illusory content and illogical form of bizarre delusions (Kavanau 1999; Charlton 2000; Charlton and Kavanau 2002).

THE SOCIAL BRAIN AND DELUSIONS

At least two types of delusion may therefore be distinguished: ToM delusions and bizarre delusions. Both types are false beliefs, and both types of false belief are typically resistant to short term and unstructured attempts at modifying them. Only the form and content of ToM type of delusions are related to the social evolutionary history of humans. By contrast, the form and content of bizarre delusions are products of pathology rather than adaptation. This illustrates the way in which evolutionary biology must be applied to medicine only with caution, because while some diseases reflect evolutionary and adaptive categories, other diseases do not.

ToM delusions may occur in multi-symptomatic syndromes such as schizophrenia, mania and psychotic depression as well as in pure cases. What makes them potentially interesting to evolutionary biologists is that ToM delusions can occur encapsulated in pure cases, where the subject is free of other psychiatric pathologies (these are the delusional disorders). In other words, ToM delusions are not necessarily pathological; their *consequences* may be pathological (e.g. a person may suffer extreme distress, may kill him/herself

or another person as a consequence of his/her false beliefs). But the cognitive nature of the delusion is simply that of normal social beliefs.

A ToM delusion is a false belief which is the *logical* outcome of a *false premise*. The processes of reasoning are intact and unimpaired, but are operating on incorrect assumptions. For 'Edward' (the jealous murderer described above), two plus two still equals four and the sun still rises in the East. Edward was also able to give accurate accounts of direct observations of objective data. However, his assumption that his girlfriend was having an affair was based upon an incorrect interpretation of indirect inferences concerning her state of mind. Although her behaviour appeared to him to be consistent with the assumption of her sexual infidelity, the fact of the matter was that Edward's girlfriend was not actually having an affair — his belief was false.

ToM delusions *happen to be* false beliefs, but they are the result of normal brain processes operating on mistaken premises concerning other people's DMIs. Mistaken premises concerning other people's DMIs are inevitable given that these inferences are based upon unreliable evidence. ToM delusions are (probably) caused by many different factors in different individuals, such as the inevitable error rate of the ToM mechanism (especially when operating under modern conditions), and by emotional abnormalities — perhaps based on personality, or perhaps due to specific pathologies affecting emotions. The fact that inferences about the DMIs of other people depend on monitoring our own emotions implies that anything which affects our body state may affect our understanding of social situations. This is a mechanism whereby changes in body chemistry, the influences of drugs, and the effects of ill-health can all lead to changes in social functioning (Charlton 2000).

The psychological mechanism leading to delusional disorders is therefore identical with the mechanism of normal human ToM. This implies that almost all religious and/or political beliefs, for instance, also have the form of ToM delusions, since they involve unwarranted assumptions about the DMIs of imagined entities, such as gods, or abstract entities, such as political parties and leaders. Indeed, *all* inferences concerning the DMIs of other people are inevitably based on incomplete and inconclusive evidence. So religious and political beliefs are not unusual, but this may explain why such culturally-dependent and evidentially-insecure beliefs nevertheless generate such powerful emotional attachments.

By contrast, bizarre delusions are the consequence of illogical thinking, can have any subject matter, and never occur in isolation. Bizarre delusions are found only as part of complex syndromes such as schizophrenia, mania, psychotic depression, dementia and delirium, conditions which include other primary psychiatric symptoms, such as hallucinations, incoherent speech and major mood changes. And bizarre delusions are caused by global brain dysfunction, such as delirium, especially when this incorporates severe and chronic sleep disturbance.

But false beliefs might also be the outcome of impaired thinking, of 'illogical' reasoning — even correct premises would not necessarily lead to correct conclusions, since thinking is impaired. When reasoning is illogical, two plus two would not necessarily equal four, but might instead equal three, or five, or Adolf Hitler. The sun may rise in the West tomorrow, fail to rise at all, or have turned into a balloon.

THERAPEUTIC IMPLICATIONS

The suggested discrimination of delusions into ToM and bizarre types is not merely an abstract exercise — the therapeutic implications are considerable. For example, when treating a person with ToM delusions, it would be a mistake to try and eradicate the false belief using drugs, since the mechanisms leading to that belief are not necessarily pathological. Drugs might usefully alter the emotional state which makes a delusion distressing, e.g. a neuroleptic such as chlorpromazine might reduce extreme fear or agitation, but the false belief would probably remain intact, even if a person held that belief with less conviction, or did not feel compelled to take any action as a consequence of that belief (Charlton 2000).

And since ToM delusions are the product of rational thought, then belief modification by rational persuasion is a possible line of therapy, and indeed 'cognitive' therapies have been applied to these kinds of belief with some success (Kingdon, Turkington and John 1994; Dolan and Bishay 1996). The main limitation of such rational persuasive therapies, however, is that it may not be possible to 'prove' the patient wrong, since the delusions may be based upon an inference concerning another person's mind (Charlton and McClelland 1999). The most fruitful approach might simply be to persuade the deluded person to acknowledge that there is *room for doubt* concerning the content of other minds, and that there are *other hypotheses* concerning the dispositions, motivations and intentions of others that are at least equally plausible, but which do not entail those beliefs which have led to that person's referral.

Changing a person's false beliefs by such techniques of persuasion would, presumably, be neither easier nor more difficult than changing a person's religious or political opinions. That is to say, such persuasion is very difficult but not impossible, and there are recognised techniques which improve the chance of success.

The treatment of bizarre delusions is a different matter altogether. At present, bizarre delusions are treated with neuroleptics, which probably do not eliminate the false belief so much as make the subject 'indifferent' to the belief (Healy 1997). But if bizarre delusions are a manifestation of delirium in patients with a history of severe sleep loss and either EEG or clinical evidence of delirium, then management directed primarily at securing deep and

restorative sleep would be expected to produce significant clinical improvement. The clinical benefit of neuroleptic and other tranquillising drugs, e.g. lorazepam, currently used in the management of acute psychosis, might turn out to be attributable mainly to their sleep-promoting effects. This topic is given more detailed consideration elsewhere (Charlton 2000).

ADAPTIVE SIGNIFICANCE OF PSYCHIATRIC PHENOMENA

There is no guarantee that the study of psychiatric phenomena will increase understanding of the social brain, but the fact that emotional and social systems are often disrupted in psychiatric illness implies that it is worth looking for a plausible evolutionary and adaptive rationale for the patterns of symptoms and signs that are observed. Delusions are only one example, and the clinical features of many other disorders appear to be a consequence of historical selection pressure.

One characteristic that may distinguish between evolved syndromes and syndromes sharing an underlying pathology may be that evolved symptoms are often characterised by a link between form and content. Where a behavioural adaptation has evolved to solve a specific problem, then adaptive behaviours will usually be associated with a specific class of stimulus (Barkow, Cosmides and Tooby 1992).

For instance, anxiety is a pathological form of an evolved, adaptive emotional state (Nesse and Williams 1996). The features of anxiety are adaptive when triggered appropriately, but in psychiatric patients the activation of the emotion may be too powerful, too sustained, activated without sufficient cause, or activated in an inappropriate situation. Among the anxiety disorders termed 'phobias', phobias of snakes and spiders are much more common than, say, phobias of lizards and ants, or of guns and automobiles. This very probably reflects the importance of snakes and spiders in human evolutionary history which has led to facilitated learning of fear to these specific categories of agent — however inappropriate the resulting behaviour may be in twenty-first century England. People are not born with innate phobias, but it is easier to learn a phobia of stimuli that represented a significant threat in our ancestral environment than to learn a phobia of novel threats (Nesse and Williams 1996).

Another example may be panic disorder, a disabling form of anxiety which is probably the adult variant of 'separation anxiety' (freezing and screaming) that is observed in infant children when they have become detached from the family group (Klein 1980; Matthews and Charlton 2000). Panic is a form of 'signalling' for rescue — almost literally a 'cry for help'. Adults with panic symptoms seem to have been more prone to separation anxiety as children, and individual cases with pure panic disorder have reported that attacks occur when away from the family base, and that panic attacks include a powerful

desire to be reunited with the family (Silove *et al*. 1996; Matthews and Charlton 2000).

In conclusion, it seems likely that psychiatry has the potential to become an important source of evidence for understanding the evolution of the social brain.

ACKNOWLEDGEMENTS

Thanks are due to Florence Walston, David Healy, Hamish McClelland and Lee Kavanau for clinical information, analysis and ideas.

REFERENCES

Antrobus, J.S. and Bertini, M. (1992). Introduction. In J.S. Antrobus, M. Bertini (eds), *The Neurobiology of Sleep and Dreaming*. Erlbaum, Hillsdale, NJ, pp. 1–16.

American Psychiatric Association (APA) (1994). *DSM-IV: Diagnostic and Statistical Manual of Mental Disorders*, 4th edn. APA, Washington, DC.

Barkow, J.H., Cosmides, L. and Tooby, J. (1992). *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford University Press, New York.

Bentall, R.P., Kinderman, P. and Kaney, S. (1994). The self, attributional processes and abnormal beliefs: towards a model of persecutory delusions. *Behav. Res. Therapeut.* **32**, 331–341.

Brüne, M. (2001). De Clérambault's syndrome (erotomania) in an evolutionary perspective. *Evol. Hum. Behav.* **22**, 409–415.

Brothers, L. (1990). The social brain: a project for integrating primate behavior and neurophysiology in a new domain. *Concepts Neurosci.* **1**, 27–51.

Buss, D.M. (1994). *The Evolution of Desire*. Basic Books, New York.

Byrne, R.W. and Whiten, A. (eds) (1988). *Machiavellian Intelligence. Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans*. Clarendon Press, Oxford.

Charlton, B.G. (1998). Psychopharmacology and the human condition. *J. R. Soc. Med.* **91**, 599–601.

Charlton, B. (2000) *Psychiatry and the Human Condition*. Radcliffe Medical Press, Oxford.

Charlton, B.G. and Kavanau, J.L. (2002). Delirium and psychosis: an integrative model. *Med. Hypotheses* **58**, 24–27.

Charlton, B.G. and McClelland, H.A. (1999). Theory of mind and the delusional disorders. *J. Nerv. Ment. Dis.* **187**, 380–383.

Damasio, A.R. (1994). *Descartes Error: Emotion, Reason and the Human Brain*. Putnam, New York.

Damasio, A.R. (1999). *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*. William Heinemann, London.

Dolan, M. and Bishay, N. (1996). The effectiveness of cognitive therapy in the treatment of non-psychotic morbid jealousy. *Br. J. Psychiat.* **168**, 588–593.

Enoch, M.D. and Trethowan, W. (1991). *Uncommon Psychiatric Syndromes*, 3rd edn. Butterworth-Heinemann, Oxford.

Everson, C.A. (1997). Clinical manifestations of sleep deprivation. In W.J. Schwartz (ed.) *Sleep Science: Integrating Basic Research and Clinical Practice*. Karger, New York, pp. 34–59.

Garety, P.A. and Hemsley, D.R. (1994). *Delusions: Investigations into the Psychology of Delusional Reasoning*. Psychology Press, Hove.

Geary, D.C., Rumsey, M., Bow-Thomas, C. C. and Hoard, M.K. (1995). Sexual jealousy as a facultative trait: evidence from the pattern of sex differences in adults from China and the United States. *Ethol. Sociobiol*. **16**, 355–383.

Healy, D. (1997). *Psychiatric Drugs Explained*. Mosby-Wolfe, London, 1997.

Humphrey, N.K. (1976). The social function of intellect. In P.P.G. Bateson and R.A. Hinde (eds). *Growing Points in Ethology*. Cambridge University Press, Cambridge.

Kavanau, J.L. (1994). Sleep and dynamic stabilization of neural circuitry: a review and synthesis. *Behav. Brain Res*. **63**, 111–126.

Kavanau, J.L. (1999). Adaptations and pathologies linked to dynamic stabilization of neural circuitry. *Neurosci. Biobehav. Rev*. **23**, 635–648.

Kingdon, D., Turkington, D. and John, C. (1994). Cognitive behavior therapy of schizophrenia. *Br. J. Psychiat*. **164**, 581–587.

Klein, D.F. (1980). Early experience with imipramine and anxiety. *Compr. Psychiat*. **12**, 411–427.

Kreuger, J.M. and Obal, F. (1993). A neuronal group theory of synaptic function. *J. Sleep Res*. **2**, 63–69.

Matthews, S., Charlton, B.G. (2000). Phenomenology of panic attacks reflects human evolutionary history of separation anxiety. *Irish Med. J*. **93**, 184–185.

Miller, G. (2000) *The Mating Mind: How Sexual Choice Shaped the Evolution of Human Nature*. Heinemann, London.

Mowat, R.R. (1966). *Morbid Jealousy and Murder*, Tavistock, London.

Mullen, P.E. and Martin, J. (1994). Jealousy: a community study. *Br. J. Psychiat*. **164**, 35–42.

Mullen, P.E. and Pathe, M. (1994). The pathological extensions of love. *Br. J. Psychiat*. **165**, 614–623.

Nesse, R.M. and Williams, G.C. (1996). *Evolution and Healing*. Phoenix, London.

Pagel, M. (1997). Desperately concealing father: a theory of parent–infant resemblance. *Anim. Behav*. **53**, 973–981.

Sheets, V.L., Fredendall, L.L. and Claypool, H.M. (1997). Jealousy evocation, partner reassurance, and relationship stability: an exploration of the potential benefits of jealousy. *Evol. Hum. Behav*. **18**, 387–402.

Shepherd, M. (1961). Morbid jealousy: some clinical and social aspects of psychiatric symptoms. *J. Ment. Sci*. **107**, 687–753.

Silove, D., Manicavasagar, V., Curtis, J., Blaszcynsia, A. (1996). Is early separation anxiety a risk factor for adult panic disorder? *Arch. Gen. Psychiat*. **37**, 475–482.

Sims, A. (1995). *Symptoms in the Mind: an Introduction to Descriptive Psychopathology*. W.B. Saunders, London.

Slaby, A.E. and Cullen, L.O. (1987). Dementia and delirium. In A. Stoudemire and B.S. Fogel (eds) *Principles of Medical Psychiatry*. Grune and Stratton, New York, pp. 135–175.

Stickgold, R. (1998). Sleep: off-line memory re-processing. *Trends Cogn. Sci*., **2**, 484–492.

Townsend, J.M. (1995). Sex without emotional involvement: an evolutionary interpretation of sex differences. *Arch. Sexual Behav*. **24**, 173–206.

Walston, F., Blennerhassett, R.C., Charlton, B.G. (2000). 'Theory of mind', persecutory delusions and the somatic marker mechanism. *Cogn. Neuropsychiat*. **5**, 161–174.

Walston, F., David, A.S. and Charlton, B.G. (1998) Sex differences in the content of persecutory delusions: a reflection of hostile threats in the ancestral environment? *Evol. Hum. Behav.* **19**, 257–260.

Whiten, A. and Byrne, R.W. (ed.) (1997). *Machiavellian Intelligence II: Extensions and Evaluations*. Cambridge University Press, Cambridge, MA.

Wiederman, M.W. and Allgeier, E.R. (1993). Gender differences in sexual jealousy: adaptionist or social learning explanation? *Ethol. Sociobiol.* **14**, 115–140.

Wilson, M. and Daly, M. (1992). The man who mistook his wife for a chattel. In J.H. Barkow, L. Cosmides and J. Tooby (eds), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. Oxford University Press, New York, pp. 289–322.

Wrangham, R. and Peterson, D. (1996). *Demonic Males: Apes and the Origins of Male Violence*. Houghton Mifflin, Boston, MA.

15

Social Cognition in Paranoia and Bipolar Affective Disorder

PETER KINDERMAN

Department of Clinical Psychology, University of Liverpool, UK

The psychological study of psychosis has undergone something of a *volte face* recently. Many of the pioneering psychoanalysts, especially Freud, studied psychosis from a primarily psychological standpoint. This approach was largely replaced, in psychiatry, with more biological perspectives, leaving psychological studies of psychotic experiences largely sidelined. More recently, however, advances in cognitive psychology, especially social cognition, have led to an upsurge of interest in both paranoia and bipolar disorder (manic depression). As well as contributing to the development of highly successful therapy programmes, this research has led to a number of interesting insights into the processes of social cognition.

FREUD AND DEFENCES

Psychoanalytic theories of paranoia and mania both commonly invoke defence mechanisms, particularly projection. Sigmund Freud suggested that delusions represent the externalisation of desires, fears or conflicts. Freud (1911/1950) proposed an extremely influential psychoanalytic theory of persecutory delusions. Paranoid delusions, Freud claimed, are consequences of a process protecting the conscious ego from awareness of conflict with unacceptable homosexual impulses stemming from the id. Such homosexual urges are, Freud argued, denied or contradicted and then countered by the defences of rationalisation and projection. A male patient is essentially confronted with

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

the idea that, 'I (a man) love him (a man)'. This is unacceptable, leading to reaction-formation and the idea: 'I do not love him, I hate him'. Such inhuman hatred is still unacceptable, and is rationalised as: 'I hate him because he hates and persecutes me'.

It is notable that Freud's model of paranoia includes two elements, a defensive projection or externalisation of threatening material, and latent homosexuality. It is possible that the defensive component is more valid than the homosexual part. Psychoanalytic writers after Freud have suggested that persecutory delusions serve a defensive function, without necessarily stressing latent homosexuality. Colby (see Winters and Neale 1983) suggested that paranoia stems from a tendency to perceive or generate threats to one's self-esteem, combined with a protective mechanism of projection and externalisation of the threat to others. One of the main benefits of Colby's theory is that it is testable. People with delusions of persecution should readily perceive potential threats to self-esteem, and they should also locate the source of such threats as external. Many more recent cognitive investigations of paranoia are compatible with Colby's model.

MODERN CONCEPTUAL FRAMEWORKS

Recent advances in psychological understanding of psychotic experiences were summarised in a report by the British Psychological Society, Division of Clinical Psychology (2000). The report outlines many of the beliefs about psychotic experiences that stem from the 'medical model' and suggests more psychological perspectives. Psychotic beliefs and experiences are more common than most people think and can be seen in healthy, well-functioning individuals. For instance, 10–15% of the normal population have had a hallucination at some point in their lives (van Os *et al.* 2000). Extreme circumstances, such as sensory or sleep deprivation, have been shown to lead to various disturbances, including paranoia and hallucinations (Hemsley 1993).

There is also substantial evidence that psychotic experiences are on a continuum with normality (Bentall, Claridge and Slade 1989; Claridge 1994; Claridge *et al.* 1996). A dimensional approach to psychotic experiences can be more useful in terms of understanding and planning care than a categorical system (van Os *et al.* 1999).

The conceptual and practical problems with the practice of diagnosis are also important. Psychiatric diagnoses are labels that describe certain types of behaviour; they do not tell us anything about the nature or causes of the experiences. If care is not taken, it may be assumed that diagnostic categories offer an explanation for unusual experiences, rather than merely a short-hand description. The central issue in diagnosis is one of classification — the idea that particular psychological problems cluster together and can therefore be

considered together. This has been termed 'carving nature at the joints' (Hamilton and Huntington 1961, p. 511), which means that it is assumed that the problems called 'schizophrenia' are different from the problems called 'bipolar disorder', in the same way that birds are different from reptiles. Many psychologists believe such distinctions are invalid, that diagnostic approaches to psychological problems do not reflect real 'joints' in nature (Bentall 1990; Boyle 1990).

Psychiatric diagnoses appear particularly unreliable. Early research (Beck *et al*. 1962; Blashfield 1973) showed that clinicians often disagreed about psychiatric diagnoses and that diagnostic practices differed from country to country. Psychiatrists have therefore put a great deal of effort into improving the consistency of diagnosis, most notably through the publication of specific manuals that specify which symptoms an individual must have for a specific diagnosis to be made. The best-known example is the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn (DSM-IV; APA 1994). However, these efforts have had only limited success in normal clinical practice (Kirk and Kutchins 1994).

Psychiatric diagnosis also lacks validity. The predictive validity of schizophrenia is very low; the outcome for people with a diagnosis of schizophrenia is extremely variable (Bleuler 1974; Ciompi 1984) and attempts to define a diagnostic group with a more uniform outcome have not been very successful (Boyle 1990). Diagnosis also appears invalid as a guide for treatment. Although neuroleptics are purported to be specific treatments for 'schizophrenia', and lithium a specific for 'bipolar disorder', the response to treatment appears to follow from the *symptoms*, not from *diagnosis* (Crowe *et al*. 1986; Naylor and Scott 1980; Moncrieff 1997). Diagnostic categories are therefore of very limited use in predicting course or outcome. Finally, statistical techniques of factor analysis (Slade and Cooper 1979) and cluster analysis (Everitt, Gourlay and Kendell 1971) have highlighted the extensive overlap between those diagnosed with schizophrenia and those diagnosed as having major affective disorder (Bentall 1990; APA 1994).

Karl Jaspers (1912/1963) followed the tradition of the great German psychiatric classifiers, Kraepelin (1896) and Bleuler (1950/1911). He was particularly interested in delusional beliefs, believing that their essential nature was the fact that they were 'ununderstandable' — that they could not be understood in terms of social or psychological processes. This claim is explicitly contradicted by modern clinical psychology. Psychological formulations, developed individually with each person, offer an alternative to diagnosis. In Britain, in particular, considerable progress has recently been achieved in understanding specific psychological mechanisms that can lead to unusual beliefs, hallucinations and difficulties in communication (Bentall 1990).

SOCIAL COGNITION

The two major practical consequences of this conceptual framework have been the development of specific therapeutic techniques and, paralleling these, the development of research strategies that address individual phenomena. Therefore, research into bipolar disorder and paranoia have advanced independently. Both, however, have advanced under the rubric of social cognition.

Richard Bentall (1990) pointed out that in order to understand delusions as phenomena with content, they need to be studied in context. Delusions are, Bentall suggested, false personal beliefs that concern 'one's place in the social universe' (Bentall 1990).

PARANOIA

Perhaps the most fundamental element of social cognition is the self-concept. Brewin (1986) reviewing the theoretical basis of cognitive-behavioural therapy, suggested that mental representations of knowledge about the self (self-concepts) underlie disorders such as depression, social phobia and generalised anxiety disorder. Markus and Wurf (1987) and Kihlstrom and Cantor (1984) outlined several lines of evidence suggesting that the self-concept directs or guides information processing, has a central role in the regulation of affect, and is implicated in a large number of interpersonal or social-cognitive processes (for reviews, see Markus, Smith and Moreland 1985; Higgins and Bargh 1987).

In paranoia, Kinderman (1994) used an 'emotional' version of the Stroop (1935) task to investigate attention to self-referent words. The emotional Stroop task assesses the degree to which words of emotional salience attract attentional resources by measuring the relative time it takes to colour-name the ink in which these words are printed. When the ink colour-naming is slow, it is taken to imply that the salience of the words themselves is causing interference. Kinderman (1994) found that, on a simple questionnaire that asked participants to endorse self-descriptive words, paranoid individuals endorsed as self-descriptive as many negative words as did 'normal' control participants, and as many negative words as did depressed participants. On the Stroop test, using these same words, paranoid patients preferentially attended to negative self-referent words. This implies that negative self-referent information, as well as threat-related material (Kaney *et al.* 1992) is pertinent to paranoid individuals.

Previous research has established that people with a diagnosis of schizophrenia have poorly elaborated (Robey, Cohen and Gara 1989) and contradictory (Gruba and Johnson 1974) self-concepts. 'Schizophrenic patients

do not have uniformly lower self-esteem than normals, but, rather, specific domains of self-esteem are affected' (Garfield, Rogoff and Steinberg 1987; p. 225).

Kinderman and Bentall (1996) examined self-actual:self-ideal discrepancies, and discrepancies between self-actual and believed parent-actual representations in paranoid, depressed and non-patient participants. They used a modified version of Higgins's (1987) Selves Questionnaire, whereby people were asked to generate words that describe their actual selves and their ideals. Kinderman and Bentall modified this approach by asking people also to describe themselves as other people see them. Normal participants showed high consistencies between all domains of the self-concept, while depressed participants showed marked self-discrepancies. Paranoid patients alone displayed a high degree of consistency between self-perceptions and self-guides, together with discrepancies between self-perceptions and the believed perceptions of parents about the self. Paranoid patients also believed that their parents had more negative views of them than did other subjects.

This curious self-concept appears to be maintained by specific causal attributions about pertinent social events. Kaney and Bentall (1989) found that patients with persecutory delusions tended to attribute hypothetical negative events to excessively external, global and stable causes, and hypothetical positive events to abnormally internal, global and stable causes when compared to relevant comparison groups, using the Attributional Style Questionnaire (ASQ; Peterson *et al*. 1982). This finding was substantially replicated by Candido and Romney (1990). This is, in itself, consistent with a defensive, externalising process.

More interestingly, these attributions appear to be observed only when the questions are obvious. Kinderman *et al*. (1992) found that independent judges typically agreed with the control subjects' self-ratings of the internality of their own causal statements on the ASQ. However, deluded subjects self-rated as external many causal statements which were rated by the independent judges as being internal. Lyon, Kaney and Bentall (1994) employed a non-obvious measure of attributional style developed by Winters and Neale (1985), which is presented to subjects as a test of memory. Essentially, people are asked to 'remember' the causes of negative and positive events, but have in fact never been given appropriate information. Both deluded and depressed patients responded similarly on this measure, by making internal attributions for negative events. However, on a traditional attributional style measure the deluded subjects made external attributions for negative events, as previously found by Kaney and Bentall (1989) and Candido and Romney (1990). Responses to this task and the Stroop task imply that paranoid individuals have an implicit set of negative self-representations, but that the externalising attributional style ('bad things are not my fault') leads to a more positive, overt, set of self-representations.

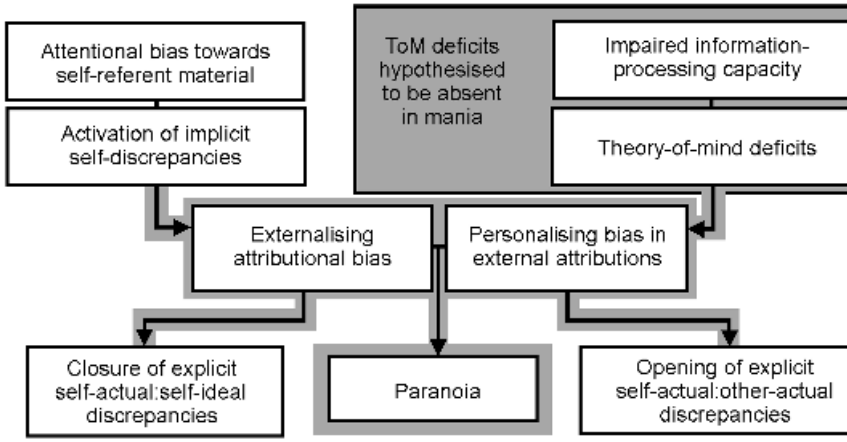


Figure 15.1. A proposed model of paranoia and mania. Adapted from Bentall and Kinderman (1999)

The story of causal attributions in paranoia gets more interesting. Recent research suggests that the bipolar internality scale of Peterson and colleagues' (1982) Attributions Style Questionnaire (most commonly used in research into causal attributions) may be in need of revision. The ASQ uses a simple dimensional internality scale — asking people to rate causes from being entirely 'due to me' to being entirely 'due to other people or circumstances'. Kinderman and Bentall (1996) have suggested a three-way categorisation of the internality dimension: internal ('due to me'), external-personal ('due to another person or other people') and external-situational ('due to the situation, circumstances or chance').

Kinderman and Bentall (1997) examined causal attributions for positive and negative hypothetical social events made by paranoid patients, depressed patients and non-patient participants using a novel measure of causal locus, the Internal, Personal and Situational Attributions Questionnaire (IPSAQ; Kinderman and Bentall 1996), designed to reflect this taxonomy. Depressed patients tended to attribute negative social events to internal (self-blaming) causes. Non-patient participants and patients with delusions of persecution tended to avoid such self-blame. However, whereas non-patient participants tended to choose situational or circumstantial external attributions, paranoid patients tended to choose external attributions that located blame in other persons.

These findings support Bentall, Kinderman and Kaney's (1994) defensive attributional model of persecutory delusions. In that model (Figure 15.1) it is hypothesised that, in deluded patients, activation of self/ideal discrepancies by threat-related information triggers defensive explanatory biases which have the

function of reducing the self/ideal discrepancies but which result in persecutory ideation. As can be seen in Figure 15.1, however, Bentall, Kinderman and Kaney hypothesised that the peculiarly paranoid, other-blaming attributions might be a result of specific problems — deficits — in theory of mind (ToM). In fact, parallel research into social cognition in bipolar disorder suggested that superficially similar attributional abnormalities might be important in that disorder also. It was initially hypothesised that ToM might be responsible for differences between the disorders.

BIPOLAR DISORDER

Bipolar disorder is a common, severe disorder characterised by recurrent episodes of depression and mania or hypomania separated by periods of relative normality. The lifetime incidence is similar to schizophrenia, at around 1% (Weissman *et al*. 1988). People who receive this diagnosis have major problems; the outcome is poor despite medication such as lithium carbonate, carbamazepine and sodium valproate (Griel *et al*. 1997) and the mean relapse rate is 50% at 1 year (see e.g. Keller *et al*. 1993) and up to 70% after 4 years (Gitlin *et al*. 1995). Between 20% and 56% of patients with a diagnosis of bipolar disorder attempt suicide during the course of their illness (Goodwin and Jamison 1990) and this is the most common diagnosis associated with the longest length of stay in acute psychiatric units (Creed *et al*. 1997).

Although mania is often thought of as the opposite to depression, this is not in fact the case. Goodwin and Jamison (1990) presented summary data from 14 studies of mood phenomenology in manic patients, concluding that irritability was more common than depression during mania, and depression ratings are sometimes higher during manic than depressive episodes (Kotin and Goodwin 1972).

In a study that echoed research into self-referent attention in paranoia, Lyon, Startup and Bentall (1999) found that both depressed and manic patients with a diagnosis of bipolar disorder demonstrated interference by depression-related words on the emotional Stroop task. Bentall, Kinderman and Manson (in press) also examined the self-concepts of patients with a diagnosis of bipolar disorder, using the approach of Higgins (1987) that had also been used in paranoia research. This study revealed an interesting pattern of results, showing both similarities and differences in the self-representations of people receiving these two diagnoses. Paranoid individuals, remember, broadly used similar words to describe themselves and their ideals. People with a diagnosis of bipolar disorder who were currently manic or hypomanic shared this pattern, although predictably people who were depressed showed self-actual:self-ideal discrepancies. However, although paranoid individuals

showed marked self-actual:other-actual discrepancies, these were not seen in manic or hypomanic individuals.

Again echoing previous research into paranoia, Lyon, Startup and Bentall (1999) used methods previously employed by Winters and Neale (1985) and Lyon and colleagues (1994) to separate explicit and implicit attributions in bipolar patients. As in the case of paranoid patients, people who were currently manic or hypomanic revealed explicit externalising attributions for negative events (compared with depressed individuals, who were self-blaming). Implicit attributions were, however, self-blaming for both groups, again revealing a negative implicit schema. At the present time, research into the attributional style of people with a diagnosis of bipolar disorder using the taxonomy of the three-way split into internal, external-personal and external-situational causes is still under way. Preliminary results (summarised in Bentall and Kinderman 1999) indicate, however, that while the attributions made by manic or hypomanic individuals are indeed external for negative events (i.e. are self-serving or defensive) they do not implicate other people, i.e. there are relatively few external-personal attributions.

In terms of both causal attributions and self-representations, therefore, people who experience manic or hypomanic episodes appear subtly different to people who become paranoid. Both groups appear to have negative implicit self-schemas, and both appear to use defensive or self-serving casual attributions. In the case of mania or hypomania, these attributions appear to be relatively benign exaggerations of normal processes (Taylor 1988) and the consequent self-schema appears positive in terms of both self-actual:self-ideal and the self-actual:other-actual consistencies. In the case of paranoia, these externalising attributions appear directly to implicate other people, and the consequent self-representations of paranoid people may be positive in respect to self-actual:self-ideal consistency, but negative with respect to marked self-actual:other-actual discrepancies.

Bentall, Kinderman and Kaney (1994) proposed a comprehensive psychological model of paranoia, which has been extended (Bentall and Kinderman 1999) to address these similarities and differences. This model is illustrated in Figure 15.1.

The obvious next question to ask is why these differences, the differences between paranoia and bipolar disorder, exist. One possibility is that external-personal attributions stem from problems in appreciating the other person's perspective and to see his or her point of view. Without taking this perspective, the offending person's behaviour can only be attributed to some kind of general disposition ('He's a *******!'). This skill is frequently termed 'theory of mind' (ToM) and has been extensively studied in the fields of autism and Asberger's syndrome (Baron-Cohen, Leslie and Frith 1985; Frith 1989; Happé and Frith 1994; Leslie 1991) as people with these conditions seem to have severe and enduring deficiencies in this domain.

THEORY-OF-MIND

In fact, ToM deficits have been studied in psychotic phenomena before. Frith (1994) has suggested that positive symptoms (hallucinations and delusions) of schizophrenia are linked to abnormalities in ToM. Recent research indicates that ToM ability may be important in this respect. Frith and colleagues (Corcoran, Mercer and Frith 1995; Frith and Corcoran 1996) have found that symptomatically paranoid but not remitted patients perform badly on ToM tasks. Our previous model of paranoid and manic social cognition therefore drew from the observation that psychotic episodes are associated with fairly severe dysfunctions of working memory and attentional capacity (Green 1992). As ToM tasks appear to make considerable demands on cognitive resources, it is possible that the ToM deficits experienced by paranoid patients reflect these more general psychological impairments. It is similarly possible that manic individuals who do not have these cognitive deficits, do not therefore have the same difficulties with ToM and hence have benign causal attributions.

Initial research findings supported this view. In a study of normal subjects, Kinderman, Dunbar and Bentall (1998) found that people who performed worst on the ToM task made the least number of external-situational attributions and the highest number of external-personal attributions.

It was also noted that there appeared to be a different mathematical relationship between complexity and error in logical as opposed to social cognition. Kinderman, Dunbar and Bentall (1998) presented people with two types of complex tasks. Both tests involved memory and described complex relationships between people (in the case of social cognition) or events (as a comparison task). In both cases, stories were read to the participants involving up to five inter-related elements. People were then asked to recall the answers to complex questions. As the complexity of the ToM questions increased, the probability of errors rose, in fact rose exponentially. A similar increase in error rate did not occur with increasing complexity of the physical, non-ToM, questions. There appears to be something unique about understanding social relationships. Kinderman, Dunbar and Bentall (1998) commented that, in the case of physical relationships, complex relationships can be recalled with relative ease because a linear skein of linkage can be drawn through the events. This is best illustrated with causal relationships, where A causes B, B causes C, C causes D, D causes E, etc. In the case of human social relationships, questions involving ToM cannot be so linked. In social relationships, person A has beliefs about persons B, C, D and E. Moreover, persons B, C, D and E all have beliefs about A, and about each other. Finally, person A has beliefs about B, and about the beliefs B has about A (and possibly even beliefs about the beliefs that B has about the beliefs that A has). These are multiplied by the

number of protagonists, and cannot be reduced to a simple linear chain. Kinderman, Dunbar and Bentall (1998) suggested that ToM problems require the simultaneous processing of an exponentially large number of variables, a complexity that could easily suffer during periods of stress.

More recently, however, a study has been published that alters the most parsimonious model of psychotic thought. It now appears quite unlikely that paranoia reflects a simple 'defence plus deficit' whereas mania reflects a 'pure' defence.

Blackshaw *et al.* (2001) investigated the self-concept, causal attributions and ToM of people with a diagnosis of Asperger's syndrome. Asperger's syndrome offers a natural experiment to test the specifics of the model outlined above. People with Asperger's syndrome have, by definition, difficulties with social cognition — specifically ToM. If the model of paranoid attributions stemming from deficient ToM is correct, they should certainly be apparent in Asperger's syndrome.

Along with Higgins' (1987) measure of self-representations, the Internal, Personal and Situational Attributions Questionnaire (IPSAQ: Kinderman and Bentall 1996) and Fenigstein and Vanable's (1992) measure of paranoia, participants with a diagnosis of Asperger's syndrome were given a novel measure of ToM. This consisted of four line drawings (see Figure 15.2), and participants were asked simply to describe what was occurring. Participants' responses were scored in terms of the number of times references were made (both spontaneously and in response to a cue question asking specifically for the thoughts and feelings of the participants). As expected (and as previous research has established (Hare 1997); paranoid anxieties (although not frank delusional beliefs) were apparent in the people with a diagnosis of Asperger's syndrome. Also predictable were the clear deficits in ToM (although people with a diagnosis of Asperger's syndrome did not differ from the comparison group after being specifically cued to refer to thoughts and feelings). Of most interest was the fact that the two groups did not differ in terms of self-concept and causal attributions. To be clear, people with a diagnosis of Asperger's syndrome did indeed lack ToM, but did not show consequent paranoid, other-blaming, attributions.

CONCLUSIONS

These findings clearly damage our original, now apparently simplistic, model of paranoid thought and mania. It clearly cannot be correct to say that paranoid, other-blaming, attributions follow from ToM deficits. In fact, even more recent research has undermined this idea more directly. Patients with a diagnosis of bipolar disorder who were in either a depressive or manic episode

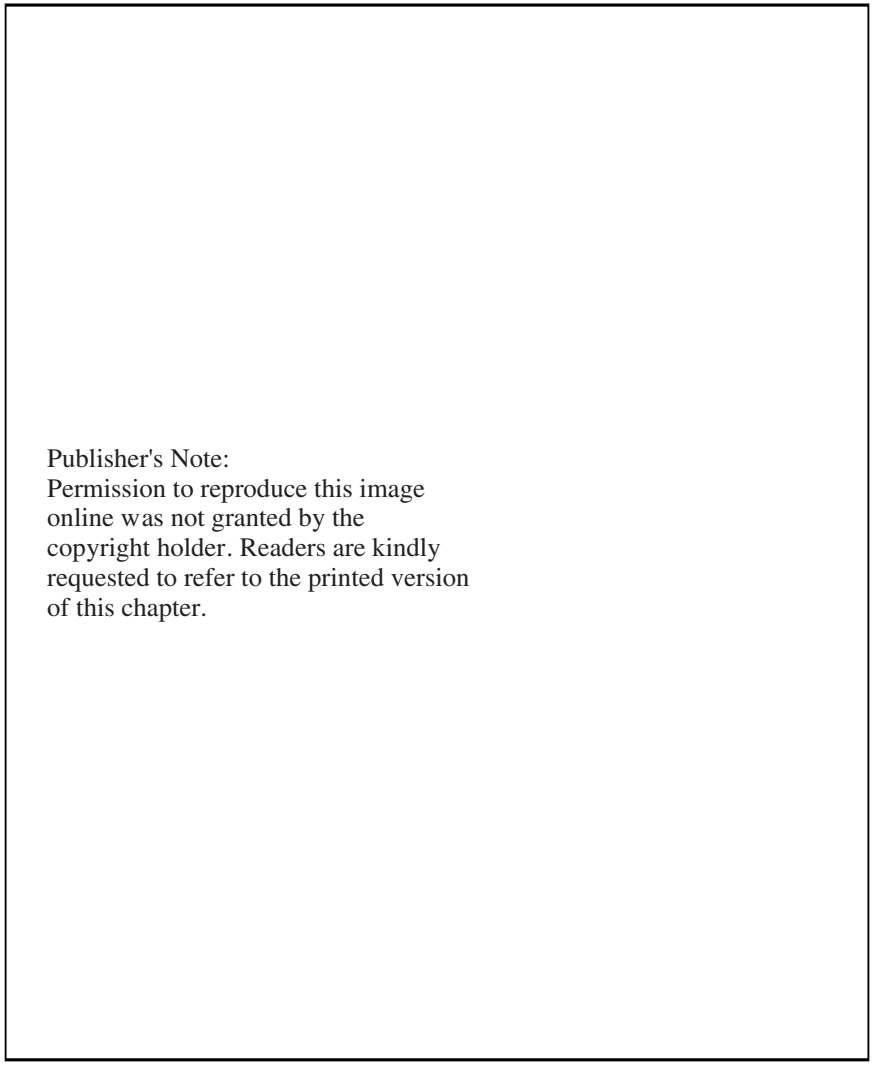


Figure 15.2. Sample images used in the assessment of theory of mind. Reproduced by permission of Sage Publications Ltd from Blackshaw *et al.* (2001)

have been seen also to exhibit ToM problems (Kerr, Dunbar and Bentall submitted).

So, what can we conclude? The following story emerges. Causal attributions are strongly implicated in the aetiology of paranoia. Paranoid individuals appear to maintain: 'I am not responsible for the bad things that are happening to me, other people are'. These causal attributions reduce self-actual:self-ideal

discrepancies but lead to the belief that others have negative views of the self. In mania, apparently similar attributions are seen, but are specifically not 'other-blaming'.

Given the fact that deficit models of ToM cannot fully explain these phenomena, we must look at the schematic nature of social cognition. It appears that, particularly when stressed, people start to have problems with ToM — they lose empathic skills and begin to find the actions of other people confusing. It looks strongly as though people who are prone to paranoia and people who are prone to mania both have negative implicit self-schemas and both use defensive, self-enhancing attributional styles. It does not look as though deficit states alone can explain why paranoid people choose paranoid explanations, while manic people do not. We might hypothesise, however, that these differences relate to the experiences of the individuals that have shaped their schemata of social cognition.

REFERENCES

APA (American Psychiatric Association) (1994). *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn. American Psychiatric Association, Washington DC.

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

Beck, A.T., Ward, C.H., Mendelson, M., Mock, J.E. and Erbaugh, J.K. (1962). Reliability of psychiatric diagnosis: II. A study of consistency of clinical judgements and ratings. *Am J. Psychiat.* **119**, 351–357.

Bentall R.P. and Kinderman P. (1999). Self-regulation, affect and psychosis: the role of social cognition in paranoia and mania. In T. Dalgleish and M. Power (eds), *Handbook of Cognition and Emotion*, 2nd edn. Wiley, Chichester, pp. 353–381.

Bentall, R.P. (ed.) (1990). *Reconstructing Schizophrenia*. Routledge, London.

Bentall, R.P., Claridge, G.S. and Slade, P.D. (1989). The multidimensional nature of schizotypal traits: a factor-analytic study with normal subjects. *Br. J. Clin. Psychol.* **28**, 363–375.

Bentall, R.P., Kinderman, P. and Kaney, S. (1994). The self, attributional processes and abnormal beliefs: Towards a model of persecutory delusions. *Behav. Res. Ther.* **32**, 331–341.

Bentall, R.P., Kinderman, P. and Manson, K. (in press). Self-discrepancies in bipolar-affective disorder. *Cogn. Ther. Res.*

Bentall, R.P. (1990). The syndromes and symptoms of psychosis. In R.P. Bentall (ed.), *Reconstructing Schizophrenia*. Routledge, London, pp. 23–60.

Blackshaw, A., Kinderman P., Hare, D. and Hatton, C. (2001). Theory-of-mind, causal attributions and paranoia in Asperger's syndrome. *Autism* **5**, 147–163.

Blashfield, R.K. (1973). An evaluation of the DSM-II classification of schizophrenia as a nomenclature. *J. Abnorm. Psychol.* **82**, 382–389.

Bleuler, E. (1950). *Dementia Praecox, or the Group of Schizophrenias* (originally published 1911; English translation, E. Zinkin). International Universities Press, New York.

Bleuler, M. (1974). The long-term course of schizophrenic psychosis. *Psychological Medicine* **4**, 244–254.

Boyle, M. (1990). *Schizophrenia: A Scientific Delusion*? Routledge, London.
Brewin, C.R. (1986). Internal attribution and self-esteem in depression: a theoretical note. *Cogn. Ther. Res.* **10**, 469–475.
British Psychological Society, Division of Clinical Psychology (2000). *Understanding Mental Illness and Psychotic Experiences: a Report by the British Psychological Society Division of Clinical Psychology*. British Psychological Society, Leicester.
Candido, C.L. and Romney, D.M. (1990). Attributional style in paranoid vs. depressed patients. *Br. J. Med. Psychol.* **63**, 355–363.
Ciompi, I. (1984). Is there really a schizophrenia? The long-term course of psychotic phenomena. *Br. J. Psychiat.* **145**, 636–640.
Claridge, G., McCreery, C., Mason, O., Bentall, R. *et al.* (1996) The factor structure of 'schizotypal' traits: a large replication study. *Br. J. Clin. Psychol.* **35**, 103–115.
Claridge, G.S. (1994) Single indicator of risk for schizophrenia: probable fact or likely myth? *Schizophr. Bull.* **20**, 151–168.
Corcoran, R., Mercer, G. and Frith, C.D. (1995). Schizophrenia, symptomatology and social inference: investigating 'theory of mind' in people with schizophrenia. *Schizophr. Res.* **17**, 5–13.
Creed, F., Tomerson, B., Anthony, P. and Tarrier, M. (1997). Predicting length of stay in psychiatry. *Psychol. Med.* **27**, 961–966.
Crowe, T.J., MacMillan, J.F., Johnson, A.I. and Johnson, E.C. (1986). The Northwick Park study of first episodes of schizophrenia: II. A controlled trial of prophylactic neuroleptic treatment. *Br. J. Psychiat.* **148**, 120–127.
Everitt, B.S., Gourlay, A.J. and Kendell, R.E. (1971). An attempt at validation of traditional psychiatric syndromes by cluster analysis. *Br J. Psychiat.* **119**, 399–412.
Fenigstein, A. and Vanable, P.A. (1992). Paranoia and self-consciousness. *J. Pers. Soc. Psychol.* **62**, 129–134.
Freud, S. (1911/1950). Psychoanalytic notes upon an autobiographical account of a case of paranoia (dementia paranoides). In J. Strachey and A. Freud (eds), *The Standard Edition of the Complete Psychological works of Sigmund Freud*, vol. III. Hogarth, London, pp. 387–466.
Frith, C. (1994). Theory of mind in schizophrenia. In A.S. David and J.C. Cutting (eds), *The Neuropsychology of Schizophrenia*. Erlbaum, Hove, pp. 147–161.
Frith, C.D. and Corcoran, R. (1996). Exploring 'theory of mind' in people with schizophrenia. *Psychol. Med.* **26**, 521–530.
Frith, U. (1989). *Autism: Explaining the Enigma*. Blackwell, Oxford.
Garfield, D.A., Rogoff, M.L. and Steinberg, S. (1987). Affect recognition and self-esteem in schizophrenia. *Psychopathology* **20**, 225–233.
Gitlin, M.J., Swenden, J., Heller, T.L. and Hammen, C. (1995). Relapse and impairment in bipolar disorder. *Am. J. Psychiat.* **152**, 1635–1640.
Goodwin, F.K. and Jamison, K.R. (1990). *Manic-depressive Illness*. Oxford University Press, Oxford.
Green, M.F. (1992). Information processing in schizophrenia. In D.J. Kavanagh (ed.), *Schizophrenia: an Overview and Practical Handbook*. Chapman and Hall, London, pp. 45–58.
Griel, W., Ludwig-Mayerhofer, W., Erazo, N., Schochlin, C. *et al.* (1997). Lithium vs. carbamazepine in the maintenance treatment of bipolar disorders: a randomised study. *J. Affect. Disord.* **43**, 151–161.
Gruba, F.P. and Johnson, J.E. (1974). Contradictions within the self-concepts of schizophrenics. *J. Clin. Psychol.* **30**, 253–254.
Hamilton, E. and Huntington, C. (1961). *The Collected Dialogues of Plato*. Princeton University Press, Princeton, NJ, p. 511.

Happé, F. and Frith, U. (1994). Theory of mind in autism. In E. Schopler and G.B. Mesibov (eds), *Learning and Cognition in Autism*. Plenum, New York, pp. 177–197.
Hare, D.J. (1997). The use of cognitive behavioural therapy with people with Asperger's syndrome: a case study. Autism. *Int. J. Res. Pract.* **1**, 215–225.
Hemsley, D.R. (1993). A simple (or simplistic?) cognitive model for schizophrenia. *Behav. Res. Ther.* **7**, 633–645.
Higgins, E.T. (1987). Self-discrepancy: a theory relating self and affect. *Psychol. Rev.* **94**, 319–340.
Higgins, E.T. and Bargh, J.A. (1987). Social cognition and social perception. *Ann. Rev. Psychol.* **38**, 369–425.
Jaspers, K. (1912/1963). *General Psychopathology* (J. Hoenig and M.W. Hamilton, trans.). Manchester University Press, Manchester.
Kaney, S. and Bentall, R.P. (1989). Persecutory delusions and attributional style. *Br. J. Med. Psychol.* **62** 191–198.
Kaney, S., Wolfenden, M., Dewey, M.E. and Bentall, R.P. (1992). Persecutory delusions and the recall of threatening and non-threatening propositions. *Br. J. Clin. Psychol.* **31**, 85–87.
Keller, M.B., Lavori, P.W., Coryell, W., Endicott, J. *et al.* (1993). Bipolar I: a five year prospective follow-up. *J. Nerv. Ment. Disord.* **181**, 238–245.
Kerr, N., Dunbar, R.I.M. and Bentall, R.P. (submitted). Theory of mind in bipolar affective disorder. *Psychol. Med.*
Kihlstrom, J.F. and Cantor, N. (1984). Mental representations of the self. In L. Berkowitz (ed.), *Advances in Experimental Social Psychology*, vol. 17. Academic Press, London, pp. 1–47.
Kinderman, P., Dunbar, R. and Bentall, R.P. (1998). Theory of mind deficits and causal attribution. *Br. J. Psychol.* **89**, 191–204.
Kinderman, P. (1994). Attentional bias, persecutory delusions and the self concept. *Br. J. Med. Psychol.* **67**, 53–66.
Kinderman, P. and Bentall, R.P. (1996). A new measure of causal locus: the Internal, Personal and Situational Attributions Questionnaire. *Personality Indiv. Diff.* **20** (2), 261–264.
Kinderman, P. and Bentall, R.P. (1996). Self-discrepancies and persecutory delusions: evidence for a defensive model of paranoid ideation. *J. Abnorm. Psychol.* **106**, 106–114.
Kinderman, P. and Bentall, R.P. (1997). Causal attributions in paranoia and depression: internal, personal and situational attributions for negative events. *J. Abnorm. Psychol.* **106** (2), 341–345.
Kinderman, P., Kaney, S., Morley, S. and Bentall, R.P. (1992). Paranoia and the defensive attributional style: deluded and depressed patients' attributions about their own attributions. *Br. J. Med. Psychol.* **65**, 371–383.
Kirk, S. and Kutchins, H. (1994). The myth of the reliability of the DSM. *J. Mind Behav.* **15**, 71–86.
Kotin, J. and Goodwin, F.K. (1972). Depression during mania. *Am. J. Psychiat.* **129**, 679–686.
Kraepelin, E. (1896). *Lehrbuch der Psychiatrie*, 5th edn. Barth, Leipzig.
Leslie, A.M. (1991). The theory of mind impairment in autism: evidence for a modular mechanism of development? In A. Whiten (ed.), *Natural Theories of Mind*. Blackwell, Oxford, pp. 63–78.
Lyon, H.M., Kaney, S. and Bentall, R.P. (1994). The defensive function of persecutory delusions: evidence from attribution tasks. *Br. J. Psychiat.* **164**, 637–646.

Lyon, H.M., Startup, M. and Bentall, R.P. (1999). Social cognition and the manic defense: attributions, selective attention, and self-schema in bipolar affective disorder. *J. Abnorm. Psychol.* **108**, 273–282.

Markus, H. and Wurf, E. (1987). The dynamic self-concept: a social psychological perspective. *Ann. Rev. Psychol.* **38**, 299–337.

Markus, H., Smith, J. and Moreland, R.L. (1985). Role of the self-concept in the perception of others. *J. Pers. Soc. Psychol.* **49**, 1494–1512.

Moncrieff, J. (1997). Lithium: evidence reconsidered. *Br. J. Psychiat.* **171**, 113–119.

Naylor, G.J. and Scott, C.R. (1980). Depot injections for affective disorders. *Br. J. Psychiat.* **136**, 105.

Peterson, C., Semmel, A., Von Baeyer, C., Abramson, L. *et al.* (1982). The Attributional Style Questionnaire. *Cogn. Ther. Res.* **3**, 287–300.

Robey, K.L., Cohen, B.D. and Gara, M.A. (1989). Self-structure in schizophrenia. *J. Abnorm. Psychol.* **98**, 436–442.

Slade, P.D. and Cooper, R. (1979). Some conceptual difficulties with the term 'schizophrenia': an alternative model. *Br. J. Soc. Clin. Psychol.* **18**, 309–317.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* **18**, 643–662.

Taylor, S.E. (1988). *Positive Illusions*. Basic Books, New York.

van Os, J., Gilvarry, C., Bale, R., van Horn, E. *et al.* (1999). A comparison of the utility of dimensional and categorical representations of psychosis. *Psychol. Med.* **29**, 595–606.

van Os, J., Hansen, M., Bijl, R. and Ravelli, A (2000). Strauss (1969) revisited: psychosis continuum in the general population? *Schizophr. Res.* **45**, 11–20.

Weissman, M.M., Leaf, P.J., Tischler, G.L., Blazer, D.G. *et al.* (1988). Affective disorders in five United States communities. *Psychol. Med.* **18**, 141–153.

Winters, K.C. and Neale, J.M. (1983). Delusions and delusional thinking: a review of the literature. *Clin. Psychol. Rev.* **3**, 227–253.

Winters, K.C. and Neale, J.M. (1985). Mania and low self-esteem. *J. Abnorm. Psychol.* **94**, 282–290.

16

Psychopathy, Machiavellianism and Theory of Mind

LINDA MEALEY[1,2] AND STUART KINNER[2]

[1]Psychology Department, College of St. Benedict, St. Joseph, MN, USA, and
[2]School of Psychology, University of Queensland, Brisbane, Queensland, Australia

CONCEPTUALISING PSYCHOPATHY

Throughout history and across cultures, human society has featured a small proportion of individuals characterised by callousness, manipulativeness, unreliability, cruelty and, sometimes, cold-blooded violence (McCord 1983). However, it was not until the turn of the nineteenth century that a scientific description of these individuals emerged. In his *Treatise on Insanity*, French psychiatrist Philippe Pinel (1806/1962) coined the term '*manie sans delire*' (insanity without delirium), arguing that it was possible to be insane ('*manie*') without a corresponding 'lesion of the understanding' ('*delire*'). Pinel and others of the time promulgated the notion that it was possible to behave in an irrational and deviant manner, despite intact intellectual functioning (Arrigo and Shipley 2001; Millon, Simonsen and Birket Smith 1998).

Possibly the first author to use the term 'psychopath' was Emil Kraepelin who, in *Psychiatry: a Textbook*, 5th edn (1896, cited in Millon, Simonsen and Birket Smith 1998), referred to individuals suffering from 'psychopathic states'. In the 7th edn of the text, published in 1903–1904, Kraepelin adopted the term 'psychopathic personalities'. This concept of the 'psychopathic personality' had been refined considerably by 1941, when American psychiatrist Hervey Cleckley published his classic text, *The Mask of Sanity*, in which 'psychopaths' were described as emotionally deficient and 'hiding behind a thin veneer of normalcy' — a 'mask of sanity'. According to Cleckley, psychopaths

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

are characterised by a kind of 'semantic aphasia', such that the emotional significance of events is lost on them — they 'know the words but not the music' (Johns and Quay 1962). Cleckley also provided a set of personality-based criteria by which psychopaths could be identified and described (see Table 16.1).

When the American Psychiatric Association released its *Diagnostic and Statistical Manual of Mental Disorders*, 1st edn. (DSM-I; APA 1952), psychopathy was referred to as 'sociopathic personality', and was defined along the lines of Cleckley's criteria. Similarly, the DSM-II (APA 1968) described what was by then called 'antisocial personality' largely in terms of inferred personality features, consistent with Cleckley's clinical conceptualisation of the psychopath.

In 1980, with the release of the third edition of the manual (DSM-III; APA 1980), the term 'antisocial personality disorder' (ASPD) was introduced. With this change in label also came shifts in diagnostic focus, from generally deviant and irresponsible behaviour to specifically criminal and antisocial conduct, and from inferred personality traits to explicit behavioural criteria. The introduction of a quantitative scoring system, too, reflected a paradigm shift among psychiatrists and psychologists, away from the notion that personality could be conceived in terms of *types*, toward the notion that personality variation was best conceived in terms of a variety of *dimensions* or *continua* of behavioural and personality *traits*. These changes in emphasis and assumptions have persisted through the most recent version of the manual (DSM-IV-TR; APA 2000)[1].

The principal reason for the changes in the most recent editions of the DSM was to increase the reliability of diagnosis, as clearly defined behavioural criteria are more easily agreed upon by clinicians than are inferred personality traits, and quantifiable measures are, generally speaking, more objective than qualitative assessments (Lilienfeld 1994; Widiger *et al.* 1996; Arrigo and Shipley 2001). However, the increase in reliability that these changes afforded was offset by a corresponding decrease in validity, specifically in discriminant validity: a much larger and more heterogeneous group of individuals receives the ASPD diagnosis than the psychopathic personality diagnosis (Hare 1996; Herpertz and Sass 2000; Abbott 2001).

The growing reliance upon dimensional measures to assess personality may mask the existence of discrete personality classes or types that underlie the normally distributed scores inherent in such measures (e.g. Gangestad and

[1] The assumption of dimensionality rather than typology has import not only for the diagnosis of ASPD, but for the entire DSM. A series of lively and significant debates has developed in an attempt to address questions concerning the functionality, practicality and veridicality of diagnoses based on the two systems (see Lilienfeld and Marino 1999; Sadler 1999; Wakefield 1992, 1999, 2000; Widiger and Clark 2000; Widiger and Sankis 2000; Zachar 2000).

Table 16.1. Cleckley's (1941/1988) psychopathy criteria

1.	Superficial charm and good 'intelligence'
2.	Absence of delusions and other signs of irrational thinking
3.	Absence of 'nervousness' or psychoneurotic manifestations
4.	Unreliability
5.	Untruthfulness and insincerity
6.	Lack of remorse and shame
7.	Inadequately motivated antisocial behaviour
8.	Poor judgement and failure to learn by experience
9.	Pathologic egocentricity and incapacity for love
10.	General poverty in major affective reactions
11.	Specific loss of insight
12.	Unresponsiveness in general interpersonal relations
13.	Fantastic and uninviting behaviour with drink and sometimes without
14.	Suicide rarely carried out
15.	Sex life impersonal, trivial, and poorly integrated
16.	Failure to follow any life plan

Snyder 1985; Meehl 1992, 1995). Further, as implications for diagnosis, treatment and prevention are different, depending on whether one uses a multidimensional model or a discrete typology model (Shipley and Arrigo 2001; Kinner 2003), most researchers (and a growing number of clinicians) prefer to assess psychopathy rather than, or in addition to, ASPD (Gacono, Loving and Bodholdt 2001; Reid 2001). In doing so, psychopathy is typically defined using scores on the Hare Psychopathy Checklist Revised or PCL-R (Hare 1991) — a measure consistent with Cleckley's original typological notion of the psychopathic personality. Comprising two correlated factors (see Table 16.2), the PCL-R measures both an interpersonal/affective component (Factor 1) and a socially deviant behavioural component (Factor 2) of psychopathy[2]. Using the PCL-R, psychopaths can be tentatively identified as those whose score falls above a designated threshold or, more in line with the typological approach, the continuum of scores generated can be conceived to reflect the probability that a given person is a psychopath.

Scores on the PCL-R are strongly predictive of general recidivism and of violence in a wide range of populations, including incarcerated offenders, forensic, and civil psychiatric patients (e.g. Hare 1991; Harris, Rice and Cormier 1991; Quinsey, Rice and Harris 1995; Rice and Harris 1995; Hill, Rogers and Bickford 1996; Salekin, Rogers and Sewell 1996; Rice 1997;

[2] Cooke and Michie (2001) argue that there are, in fact, three factors underlying psychopathy: arrogant and deceitful interpersonal style; deficient affective experience; and impulsive and irresponsible behavioural style. The first two of these discriminate different elements that appear in Factor 1 of the PCL-R, while the latter coincides well with Factor 2 of the PCL-R.

Table 16.2. Hare's (1991): psychopathy checklist, revised

Factor 1. Callous and remorseless use of others
- Glibness/superficial charm
- Grandiose sense of self-worth
- Pathological lying
- Conning/manipulative
- Lack of remorse or guilt
- Shallow affect
- Callous/lack of empathy
- Failure to accept responsibility for own actions

Factor 2: Chronically unstable and antisocial lifestyle
- Need for stimulation/proneness to boredom
- Parasitic lifestyle
- Poor behavioural controls
- Early behavioural problems
- Lack of realistic, long-term goals
- Impulsivity
- Irresponsibility
- Juvenile delinquency
- Revocation of conditional release

Other items (not loading on either factor)
- Promiscuous sexual behaviour
- Many short-term marital relationships
- Criminal versatility

Hemphill, Hare and Wong 1998; Quinsey *et al.* 1998; Seto and Barbaree 1999; Harris, Skilling and Rice 2001). Importantly, predictions based on the PCL-R are often different from, and significantly better than, those arrived at using DSM criteria.

PSYCHOPATHY AND MACHIAVELLIANISM

Despite current support for the notion of personality continua, evidence is accumulating that a phenotype described by severe, frequent, persistent, and life-long antisocial behaviour may indeed be a manifestation of a discrete typology (Moffitt 1993; Harris, Rice and Quinsey 1994; Ayers 2000; Lalumiere, Harris and Rice 2001; Moffitt and Caspi 2001; Skilling, Quinsey and Craig 2001). The existence of such a psychopathic 'type' would fit not only with the traditional conceptualisation of the psychopath and with the empirically documented non-reciprocal overlap between psychopathy and other measures of antisociality (e.g. Blackburn 1975, 1988; Eysenck 1977, 1987, 1998; Lalumiere and Quinsey 1996; Patrick, Zempolich and Levenston 1997;

Darke, Kaye and Finlay-Jones 1998; Dyce and O'Connor 1998; Widiger 1998; Widiger and Lynam 1998), but also with predictions from comparative and evolutionary psychology.

From an evolutionary perspective, there are good reasons to believe that psychopathy may be a discrete personality type (Mealey 1995, 1997). Mathematical modelling has demonstrated that minority, even seemingly-maladaptive, phenotypes can be maintained in populations through frequency-dependent selection, and in many animal species, one or more infrequent phenotypes do, in fact, coexist alongside the numerically dominant phenotype. Minority phenotypes in other species are most evident in visible features, such as colour morphs, but Budaev (1998), Gosling (2001), Wilson (1998) and others have also documented heritable personality differences, including aggression and other psychopathy-like attributes, in vertebrates from fish to primates (Clarke and Boinski 1995; King and Figueredo 1997; Lilienfeld *et al*. 1999; O'Connor *et al*. 2000; Weiss, King and Figueredo 2000).

Most relevant to psychopathy is the evolution of what are referred to in non-human species as 'sneaker-' or 'cheater-morphs' (e.g. Bass 1992; Gross 1996; Simmons, Tomkins and Hunt 1999). 'Sneakers' and 'cheaters' rely on non-normative, non-cooperative strategies for accruing resources and/or reproductive opportunities. As the label implies, 'sneakers' usually maintain a low profile and use deception; the category 'cheaters' includes 'sneakers', but also includes individuals that use strategies relying on force, such as theft or rape, and other forms of social manipulation. Several authors have suggested that psychopaths may be the human equivalent of animal 'cheaters' (MacMillan and Kofoed 1984; Harpending and Sobus 1987; Frank 1988; Dugatkin 1992; Mealey 1995; Lalumiere and Quinsey 1996; Colman and Wilson 1997; Seto *et al*. 1997), i.e. that psychopaths are 'designed' by natural selection to be specialised morphs that are highly effective at accruing resources and reproductive opportunities through deception, force and social manipulation.

Social psychologists have already developed a literature related to self-serving and manipulative interpersonal behaviour. Named after the infamous sixteenth century Italian author Niccolò Machiavelli, the 'Machiavellian' personality has been conceptualised in a self-report measure known as the Mach-IV (Christie and Geis 1970). 'High Machs' (those who score high on the Mach-IV) are, like psychopaths, exploitative, calculating and deceitful; they also view others as weak, untrustworthy and self-serving (Fehr, Samson and Paulhus 1992). High Machs have been found to be dominant, hostile, authoritarian and emotionally detached, yet higher in trait anxiety than their peers. On an interpersonal level, High Machs are manipulative and persuasive, but they are themselves less easily persuaded than others. High Machs are also described as being 'morally flexible' (Christie and Geis 1970; Geis 1978; Fehr, Samson and Paulhus 1992).

Like psychopaths, High Machs score high on the P (psychoticism) and E (extraversion) dimensions of the Eysenck Personality Questionnaire (Allsopp, Eysenck and Eysenck 1991; Harpur, Hare and Hakstian 1989), high on the Pd (psychopathic deviate) scale of the MMPI (Smith and Griffith 1978), and high on the more maladaptive subscales of the Narcissistic Personality Inventory and Personality Diagnostic Questionnaire (McHoskey 1995, 2001). Also like psychopaths, they are more likely to be male, and more likely to exhibit a variety of self-serving, deceitful and coercive sexual tactics (McHoskey 2001b; Seto *et al.* 1997; Seto and Lalumiere 2000). Most importantly, there is a significant correlation between Mach-IV scores and PCL-R scores, with the covariance being greatest with items loading on Factor 1, the personality component of the PCL-R (Hare 1991; Widiger *et al.* 1996)[3].

In many respects, High Machs resemble the prototypical psychopath (Fehr, Samson and Paulhus 1992; McHoskey, Worzel and Szyarto 1998; Smith 1999). As with the ASPD–psychopathy relationship, however, not all High Machs are expected to be identified as psychopaths, even though one might expect all psychopaths to be High Machs (Smith 1999; Kinner, Mealey and Slaughter 2001; see Figure 16.1). Perhaps psychopaths are those 'High Machs who have run up against the law' (Fehr, Samson and Paulhus 1992, p. 87) or, from the corollary perspective, perhaps High Machs are those psychopaths who have found socially acceptable means of meeting their ego and other needs. Cleckley believed in the commonness of such 'successful' psychopaths, providing examples of psychopathic businessmen, scientists, psychiatrists and physicians (Cleckley 1941/1988). Although in the minority, several contemporary theorists hold this belief as well (e.g. Babiak 1995a,b; Bailey 1995; Mealey 1995; McHoskey, Worzel and Szyarto 1998; but see Harris, Skilling and Rice 2001).

PSYCHOPATHY, EMPATHY AND THEORY OF MIND (ToM)

Underlying the psychopath's instrumental and Machiavellian 'cold-heartedness' is a fundamental and profound lack of empathy (Mealey and Kinner in press). Yet despite that this observation is nothing new, the nature of the apparent deficit is still unclear and extremely controversial. Most authors would suggest that the psychopath's lack of empathy results from some form of developmental pathology (e.g. Chandler and Moran 1990; Rygaard 1998; McCord 2001). However, if psychopaths are human cheater-morphs designed by natural selection, their lack of empathy should not be considered to be a 'deficit' at all, but instead, a feature of their design — their different human

[3] Kinner, Mealey and Slaughter (2001) found an even higher correlation between Mach-IV scores and the self-report version of the PCL (Hare 1990). This version of the PCL, however, does not have a factor structure.

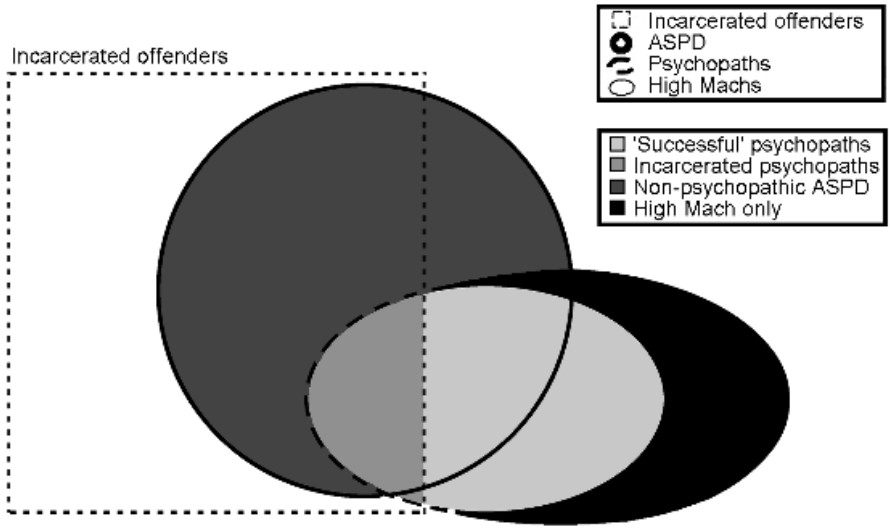


Figure 16.1. Relationships among psychopathy, Machiavellianism and ASPD

'nature' (Mealey 1997; Harris, Rice and Lalumiere 2001; Lalumiere, Harris and Rice 2001).

Empathy requires the ability to simulate the emotional state of another individual (Brothers 1990), i.e. to be able to 'walk in another's shoes' or 'get inside another's skin'. Therefore, the more similar any two individuals might be, the more similar will be their physical and mental experience, and the better will be their 'simulations' and their ability to empathise with one another (Preston and de Waal in press). But because psychopaths have a truly different design — a different nature — they are unable to accurately simulate the emotional experiences of others and, therefore, are unable to empathise with them (Mealey 1997; Mealey and Kinner in press).

Specifically, psychopaths exhibit an underarousal of what Gray (1982, 1987) calls the Behavioural Inhibition System. Psychopaths are relatively insensitive to low levels of stimulation (Gray 1987; Newman and Wallace 1993), and they do not exhibit typical autonomic or somatic responses to situations and stimuli which normally elicit anxiety or fear in others (Lykken 1957, 1995; Eysenck and Gudjonsson 1989; Williamson, Harpur and Hare 1991; Patrick, Bradley and Lang 1993; Patrick, Zempolich and Levenston 1997; Herpertz *et al*. 2001). This different physiology not only explains the psychopath's impulsivity, sensation seeking and poor passive avoidance learning (Zuckerman, Buchsbaum and Murphy 1980; Ellis 1987; Newman and Wallace 1993; Lykken 1995; Lalumiere and Quinsey 1996; Daderman and af Klinteberg 1997; Newman

1998; Blair 1999; Daderman 1999; Herpertz and Sass 2000), but also renders him/her unable to experience the full range of emotions that most humans naturally do. If it is true that 'information about the self is used to model the states of others' (Gallup 1998), then psychopaths will never be capable of fully empathising with others because their own physical and phenomenological self is, in fact, quite different from that of others.

Some theorists have argued that the psychopath's inability to empathise with others is due to a missing or undeveloped ToM module (Blair *et al*. 1995; Blair in press). Yet when Blair *et al*. (1996) compared the ToM ability of 25 psychopaths and 25 incarcerated controls, the two groups did not differ and all participants performed within the normal range (see also Happé and Frith 1996). We argue that psychopaths do not lack a ToM module, neither do they have an underdeveloped ToM module; rather, we suggest that, by virtue of their radically different phenomenological experience, the inputs that go into a psychopath's ToM module produce an output (a simulation) that is utterly unlike the actual experience of their partner (or combatant, or victim).

On the other hand, although the emotional inputs to a psychopath's ToM module are different from those actually experienced by the others with whom they interact, there is no reason to believe that the psychopath's non-emotional, i.e. cognitive, inputs will be different. Although psychopaths cannot 'feel' what others feel phenomenologically and idiographically, we argue that they can and do learn how to 'read minds' through nomothetic and actuarial analysis of their own and others' behaviour (Mealey 1992, 1997; Mealey and Kinner in press). In fact, since psychopaths cannot simulate the emotional experiences of others, in order to predict others' behaviour and to 'succeed' in social interactions (i.e. obtain their desired objective), they must rely much more than most of us on such cognitive inputs. This difference in the psychopath's phenomenology is what leads to the frequently noted reliance of identified psychopaths on a 'paint by numbers' approach in order to 'learn the appropriate emotional responses to everyday events' (Hare 1993, p. 54).

It is this lack of empathic ability, in conjunction with the consequent forced reliance on conscious monitoring of the contingencies surrounding others' behaviour, that is perceived as the psychopath's 'Machiavellian cold-heartedness' (Mealey and Kinner in press). From an evolutionary perspective, the psychopath is designed in a way that allows him to develop a ToM that understands others in purely instrumental terms: unlike the rest of us, the psychopath is unencumbered by any physiological or psychological simulation of the emotional element of another's distress, suffering, attachment or sense of fair-play. This design frees the psychopath to act in a purely egocentric and selfish manner, without the constraints typically imposed by feelings of reciprocity, guilt or shame. The 'superficial charm' and occasional (seemingly) prosocial motivation of the psychopath are simply acquired techniques—

honed by years of feedback and operant conditioning — for achieving personal gain.

FUTURE DIRECTIONS

While the design of a psychopath may not allow him to fully appreciate his impact on others, that design is, in essence, one that maximises appreciation of personal gain. To socialise a psychopath, therefore, one must rely more on the carrot and less on the stick. This suggestion may at first seem utterly inappropriate, or even raise anger and resentment amongst those who have personal experience with the callous actions of an antisocial psychopath, but antisocial behaviour is not a necessary outcome of the psychopath's lack of empathy (Kinner 2003; Mealey and Kinner in press), and while the psychopathic child will not be easily socialised by traditional measures, he is not necessarily a poor learner. Because the psychopathic child is not responsive under conditions of aversion learning or social reward, this leaves us with the difficult, but truly more rational strategy of channelling the child's energies into prosocial activities through an appropriately tailored, very tangible reward schedule (Lykken 1995; Mealey 1995; Vila 1997).

Clearly, the best approach for managing these 'fledgling psychopaths' (Lynam 1996) is prevention — going further 'upstream' (Lykken 1995). But how far upstream can and should we go? With growing evidence of Machiavellian and psychopathic 'callous/unemotional' traits in children as young as 6 (Frick *et al.* 1994; Frick 1998; Repacholi *et al.* 2001), it should be possible to identify psychopaths before their teenage years — well before they embark on an antisocial life path. With sufficiently skilled, patient and persistent parenting (or alloparenting), these temperamentally difficult children can become productive (albeit controversial) members of society. At a minimum, parents need help in identifying high-risk children, and instruction in how to take a practical, assertive approach with them (Magid and McKelvie 1987; Garmezy 1991), while using a more inductive, empathic approach with their other children (Kochanska 1991, 1993; Kochanska and Murray 2000).

Unfortunately, the congenital lack of empathy, aggressivity and impulsivity of the children who need the most attention is often evident in their relatively unskilled and indifferent parents as well (Lykken 1995). This lack of 'goodness of fit' between parental style and the needs of the child is probably an important factor in the exacerbation of conduct disorder (Lee and Bates 1985; Landy and Peters 1992; Wachs 1992; Moffitt 1993). Furthermore, the cause-and-effect relationship between parental behaviour and child behaviour is not likely to be one-way: children of different temperaments respond differentially to different socialisation techniques (McCord 1983; Dienstbier 1984; Radke-Yarrow and Zahn-Waxler 1986; Lytton 1990; Kochanska 1991, 1993;

Kochanska and Murray 2000) and, to some extent, difficult children *elicit* poor parenting (Buss 1981; Lee and Bates 1985; Bell and Chapman 1986; Lytton 1990; Snyder and Patterson 1990; Eron, Huesmann and Zelli 1991; Patterson 1992; Rowe 1994; Harris 1998).

Lykken (1995, 1998, 2001) has therefore suggested an even more radical step: the introduction of parental licensure. Lykken advocates compelling prospective parents to meet some basic criteria (e.g. absence of drug dependency, financially self-sufficient) before being permitted to have children. This approach could very well be effective; however, the practical and ethical implications of its implementation make it unacceptable to many, and exceedingly controversial to the rest. Preventive measures designed to meet the thrill-seeking and ego-gratifying needs of our 'fledgling psychopaths' might have a similar positive effect, and while the need for them may be difficult for many to appreciate, they may at least seem less controversial in the comparative light generated by Lykken's proposal!

For the psychopathic adult whose antisocial behaviour is entrenched, solutions are even less obvious. Despite extensive and intensive effort, no-one has been able to successfully change ('rehabilitate') psychopaths who have gone awry of the law (Ogloff, Wong and Greenwood 1990; Hare 1993, 1998; Harris, Rice and Cormier 1994; Rice 1997; Quinsey *et al.* 1998; Seto and Barbaree 1999; Harris, Skilling and Rice 2001). Because the psychopath is incapable of true empathy, traditional modes of treatment, such as empathy training, anger control and therapeutic community, are ineffective. Worse, many 'treatments' have actually aided the intelligent, Machiavellian psychopath in his acquisition of (superficially) social skills, allowing him greater possibility of personal success at greater cost to society (Rice, Harris and Cormier 1992; Seto and Barbaree 1999). Barring the extremely remote possibility of some (far) future genetic, pharmacologic or surgical intervention, it would seem that our challenge is not to try to change what seems to be an essential psychopathic type, but to acknowledge the psychopath for who he is, and to develop social structures and arrangements that can accommodate him at minimal cost. To be successful, such interventions must convince the psychopath that antisocial behaviour is not in his own best interests (Hare 1993; Mealey 1995; Wong and Hare in press).

This does not necessarily mean prison or threat of prison. Although many older psychopaths have developed their antisocial skills to such proficiency that they cannot be motivated to invest in an entirely new repertoire, other psychopathic (and non-psychopathic) offenders might successfully be tempted to develop their (sometimes prodigious) potential in ways that are harmless or even of benefit to society. Corrections departments in some areas have been able to funnel the sensation-seeking, absence of fear, and physical and psychological intensity of some of their worst offenders into forest fire-fighting, and even wild horse training (e.g. see California Department of Corrections

2000; BLM Colorado District Office 2002; Massachusetts Department of Corrections 2002, South Dakota Department of Corrections 2002). Preliminary data from Australian inmates suggests that most have personality profiles that would mesh well with the tasks and life circumstances of park rangers, pilots, outdoor recreation facilitators, and other professions that involve extensive freedom of movement, some element of risk, pleasure or power, and a sense of personal destiny (Kinner, Mealey and Slaughter 2001). It is not impossible to imagine psychopaths as successful (and relatively harmless) stuntmen, bush pilots, actors, jackaroos or even demolitions experts.

The key to constraining the psychopath's activities to those that are essentially harmless or even prosocial lies not in motivating him with threats of punishment, which mean very little, but with promises of tangible and ego-satisfying rewards. Lacking in empathy but with an intact theory of mind, the Machiavellian psychopath cannot be changed fundamentally but perhaps, with appropriate and targeted interventions, his destructive and antisocial behaviour can.

REFERENCES

Abbott, A. (2001). Into the mind of a killer. *Nature* **410**, 296–298.

Allsopp, J., Eysenck, H.J. and Eysenck, H.B.G. (1991). Machiavellianism as a component in psychoticism and extraversion. *Pers. Individ. Diff.* **12**, 29–41.

APA (American Psychiatric Association) (1952). *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association, Washington, DC.

APA (American Psychiatric Association) (1968). *Diagnostic and Statistical Manual of Mental Disorders*, 2nd edn. American Psychiatric Association, Washington, DC.

APA (American Psychiatric Association) (1980). *Diagnostic and Statistical Manual of Mental Disorders*, 3rd edn. American Psychiatric Association, Washington, DC.

APA (American Psychiatric Association) (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn. Text Revision. American Psychiatric Association, Washington, DC.

Arrigo, B.A. and Shipley, S. (2001). The confusion over psychopathy (I): historical considerations. *Int. J. Offender Ther. Comp. Criminol.* **45**, 325–344.

Ayers, W.A. (2000). *Taxometric Analysis of Borderline and Antisocial Personality Disorders in a Drug and Alcohol Dependent Population*. Fordham University, New York.

Babiak, P. (1995a). When psychopaths go to work: a case study of an industrial psychopath. *Appl. Psychol. Int. Rev.* **44**, 171–188.

Babiak, P. (1995b). Psychopathic manipulation in organisations: pawns, patrons, and patsies. *Issues Criminol. Legal Psychol.* **24**, 12–17.

Bailey, K. (1995). The sociopath: cheater or warrior hawk? *Behav. Brain Sci.* **18**, 542–543.

Bass, A. (1992). Dimorphic male brains and alternative reproductive tactics in a vocalizing fish. *Trends Neurosci.* **15**, 139–145.

Bell, R.Q. and Chapman, M. (1986). Child effects in studies using experimental or brief longitudinal approaches to socialisation. *Dev. Psychol.* **22**, 595–603.

Blackburn, R. (1975). An empirical classification of psychopathic personality. *Br. J. Psychiat.* **127**, 456–460.
Blackburn, R. (1988). On moral judgements and personality disorders: the myth of psychopathic personality revisited. *Br. J. Psychiat.* **153**, 505–512.
Blair, R.J.R. (1999). Responsiveness to distress cues in the child with psychopathic tendencies. *Pers. Individ. Diff.* **27**, 135–145.
Blair, R.J.R. (in press). Did Cain fail to represent the thoughts of Abel before he killed him? The relationship between theory of mind and aggression. In B. Repacholi and V. Slaughter (eds), *Individual Differences in Theory of Mind: Implications for Typical and Atypical Development*. Psychology Press, Hove.
Blair, R.J.R., Sellars, C., Strickland, I., Clark, F. *et al.* (1995). Emotion attributions in the psychopath. *Personality Individ. Diff.* **19**, 431–437.
Blair, R.J.R., Sellars, C., Strickland, I., Clark, F. *et al.* (1996). Theory of mind in the psychopath. *J. Forens. Psychiat.* **7**, 15–25.
BLM Colorado District Office (2002). Colorado Wild Horse-Inmate Program (Internet site). Retrieved 27 January 2002 from http://www.kbrhorse.net/wclo/blmco05.html
Brothers, L. (1990). The neural basis of primate social communication. *Motivation Emotion* **14**, 81–91.
Budaev, S.V. (1998). How many dimensions are needed to describe temperament in animals: a factor analysis of two data sets. *Int. J. Comp. Psychol.* **11**, 17–29.
Buss, D.M. (1981). Predicting parent–child interactions from children's activity level. *Dev. Psychol.* **17**, 59–65.
California Department of Corrections (2000). California inmates do their part during busy fire season (media release). Retrieved 29 August 2000 from http://www.cdc.state.ca.us/news/2000/00-07.htm
Chandler, M. and Moran, T. (1990). Psychopathy and moral development: a comparative study of delinquent and nondelinquent youth. *Dev. Psychopathol.* **2**, 227–246.
Christie, R. and Geis, F.L. (eds) (1970). *Studies in Machiavellianism*. Academic Press, New York.
Clarke, A.S. and Boinski, S. (1995). Temperament in non-human primates. *Am. J. Primatol.* **37**, 103–125.
Cleckley, H. (1941/1988). *The Mask of Sanity*, 5th edn. Mosby, St. Louis, MO.
Colman, A.M. and Wilson, J.C. (1997). Antisocial personality disorder: an evolutionary game theory analysis. *Legal Criminol. Psychol.* **2**, 23–34.
Cooke, D.J. and Michie, C. (2001). Refining the construct of psychopathy: towards a hierarchical model. *Psychol. Assess.* **13**, 171–188.
Daderman, A. and af Klinteberg, B. (1997). Personality dimensions characterizing severely conduct disordered male juvenile delinquents. *Rep. Dept. Psychol. Univ. Stockholm* **831**, 1–21.
Daderman, A.M. (1999). Differences between severely conduct-disordered juvenile males and normal juvenile males: the study of personality traits. *Personality Individ. Diff.* **26**, 827–845.
Darke, S., Kaye, S. and Finlay-Jones, R. (1998). Antisocial personality disorder, psychopathy and injecting heroin use. *Drug Alcohol Depend.* **52**, 63–69.
Dienstbier, R.A. (1984). The role of emotion in moral socialization. In C.E. Izard, J. Kagan and R.B. Zajonc (eds), *Emotions, Cognition and Behaviour*. Cambridge University Press, Cambridge, pp. 484–514.
Dugatkin, L.A. (1992). The evolution of the 'con artist'. *Ethol. Sociobiol.* **13**, 3–18.
Dyce, J.A. and O' Connor, B.P. (1998). Personality disorders and the five-factor model: a test of facet-level predictions. *J. Pers. Disord.* **12**, 31–45.

Ellis, L. (1987). Relationships of criminality and psychopathy with eight other apparent behavioral manifestations of sub-optimal arousal. *Pers. Individ. Diff.* **8**, 905–925.
Eron, L.D., Huesmann, L.R. and Zelli, A. (1991). The role of parental variables in the learning of aggression. In D.J. Pepler and K.H. Rubin (eds), *The Development and Treatment of Aggression*. Erlbaum, Hillsdale, NJ, pp. 169–188.
Eysenck, H.J. (1977). *Crime and Personality*, 3rd edn. Routledge and Kegan Paul, London.
Eysenck, H.J. (1987). The definition of personality disorders and the criteria appropriate for their description. *J. Personal. Disord.* **1**, 211–219.
Eysenck, H.J. (1998). Personality and crime. In T. Millon and E. Simonsen (eds), *Psychopathy: Antisocial, Criminal, and Violent Behaviour*. Guilford, New York.
Eysenck, H.J. and Gudjonsson, G.H. (1989). *The Causes and Cures of Criminality*. Plenum, New York.
Fehr, B., Samson, D. and Paulhus, D.L. (1992). The construct of Machiavellianism: twenty years later. In C.D. Spielberger and J.N. Butcher (eds), *Advances in Personality Assessment*, vol. 9, Erlbaum, Hillsdale, NJ, pp. 77–116.
Frank, R.H. (1988). *Passions within Reason: The Strategic Role of the Emotions*. W.W. Norton, New York.
Frick, P.J., O'Brien, B.S., Wootton, J.M. and McBurnett, K. (1994). Psychopathy and conduct problems in children. *J. Abnorm. Psychol.* **103**, 700–707.
Frick, P.J. (1998). Callous-unemotional traits and conduct problems: applying the two-factor model of psychopathy to children. In D.J. Cooke and A.E. Forth, and R.D. Hare (eds), *Psychopathy: Theory, Research, and Implications for Society*. Kluwer, Dordrecht, pp. 161–188.
Gacono, C.B., Loving, J.L. Jr. and Bodholdt, R.H. (2001). The Rorschach and psychopathy: Toward a more accurate understanding of the research findings. *J. Personal. Assess.* **77**, 16–38.
Gallup, G.G. Jr. (1998). Self-awareness and the evolution of social intelligence. *Behav. Processes* **42**, 239–247.
Gangestad, S. and Snyder, M. (1985). 'To carve nature at its joints': on the existence of discrete classes in personality. *Psychol. Rev.* **92**, 317–349.
Garmezy, N. (1991). Resiliency and vulnerability to adverse developmental outcomes associated with poverty. *Am. Behav. Scient.* **34**, 416–430.
Geis, F.L. (1978). Machiavellianism. In H. London and J.E. Exher (eds), *Dimensions of Personality*. Wiley Interscience, New York, pp. 305–363.
Gosling, S.D. (2001). From mice to men: what can we learn about personality from animal research? *Psychol. Bull.* **127**, 45–86.
Gray, J.A. (1982). *The Neuropsychology of Anxiety: an Enquiry into the Functions of the Septo-hippocampal System*. Clarendon/Oxford University, New York.
Gray, J.A. (1987). *The Psychology of Fear and Stress*, 2nd edn. Cambridge University Press, New York.
Gross, M.R. (1996). Alternative reproductive strategies and tactics: diversity within sexes. *Trends Ecol. Evol.* **11**, 92–98.
Happé, F.G.E. and Frith, U. (1996). Theory of mind and social impairment in children with conduct disorder. *Br. J. Dev. Psychol.* **14**, 385–398.
Hare, R.D. (1990). The Self-Report Psychopathy Scale II (unpublished: used with permission). Department of Psychology, University of British Columbia, Vancouver.
Hare, R.D. (1991). *Manual for the Hare Psychopathy Checklist, Revised*. Multi-Health Systems, Toronto.
Hare, R.D. (1993). *Without Conscience: the Disturbing World of the Psychopaths among Us*. Pocket Books, New York.

Hare, R.D. (1996). Psychopathy and antisocial personality disorder: a case of diagnostic confusion. *Psychiat. Times* **13**, 39–40.
Hare, R.D. (1998). The Hare PCL-R: some issues concerning its use and misuse. *Legal Criminol. Psychol.* **3**, 99–119.
Harpending, H. and Sobus, J. (1987). Sociopathy as an adaptation. *Ethol. Sociobiol.* **8**, 63s–72s.
Harpur, T.J., Hare, R.D. and Hakstian, A.R. (1989). Two-factor conceptualization of psychopathy: construct validity and assessment implications. *Psychol. Assess.* **1**, 6–17.
Harris, J. (1998). *The Nurture Assumption: Why Children Turn Out the Way They Do*. Free Press, New York.
Harris, G.T., Rice, M.E. and Cormier, C.A. (1991). Psychopathy and violent recidivism. *Law Hum. Behav.* **15**, 625–637.
Harris, G.T., Rice, M.E. and Cormier, C.A. (1994). Psychopaths: is a therapeutic community therapeutic? *Therapeut. Commun. Int. J. Therapeu. Support. Org.* **15**, 283–299.
Harris, G.T., Rice, M.E. and Lalumiere, M. (2001). Criminal violence: the roles of psychopathy, neurodevelopmental insults, and antisocial parenting. *Crim. Just. Behav.* **28**, 402–426.
Harris, G.T., Rice, M.E. and Quinsey, V.L. (1994). Psychopathy as a taxon: evidence that psychopaths are a discrete class. *J. Consult. Clin. Psychol.* **62**, 387–397.
Harris, G.T., Skilling, T.A. and Rice, M.E. (2001). The construct of psychopathy. *Crime Just.* **28**, 197–264.
Hemphill, J.F., Hare, R.D. and Wong, S. (1998). Psychopathy and recidivism: a review. *Legal Criminol. Psychol.* **3**, 139–170.
Herpertz, S.C. and Sass, H. (2000). Emotional deficiency and psychopathy. *Behav. Sci. Law* **18**, 567–580.
Herpertz, S.C., Werth, U., Lukas, G., Quanaibi, M. *et al.* (2001). Emotion in criminal offenders with psychopathy and borderline personality disorder. *Arch. Gen. Psychiat.* **58**, 737–745.
Hill, C.D., Rogers, R. and Bickford, M.E. (1996). Predicting aggressive and socially disruptive behaviour in a maximum security forensic psychiatric hospital. *J. Forens. Sci.* **41**, 56–59.
Johns, J.H. and Quay, H.C. (1962). The effect of social reward on verbal conditioning in psychopathic and neurotic military offenders. *J. Consult. Psychol.* **26**, 217–220.
Kaye, S., Darke, S., Finlay-Jones, R. and Hall, W. (1997). *The Relationship between Antisocial Personality Disorder, Psychopathy and Injecting Heroin Use*. Monograph 32, National Drug and Alcohol Research Centre (NDARC), Sydney.
King, J.E. and Figueredo, A.J. (1997). The five-factor model plus dominance in chimpanzee personality. *J. Res. Personality* **31**, 257–271.
Kinner, S. (2003). Psychopathy as an adaptation: implications for society and social policy. In R. Bloom and N. Dees (eds), *Evolutionary Psychology and Violence: a Primer for Policymakers and Public Policy Advocates*. Praeger, New York.
Kinner, S.A., Mealey, L.J. and Slaughter, V.P. (2001). Psychopathy, Machiavellianism, risk-taking and career preferences. Paper presented at the Australian and New Zealand Association of Psychiatry, Psychology and Law (ANZAPPL) Annual Conference, Melbourne, Australia.
Kochanska, G. (1991). Socialization and temperament in the development of guilt and conscience. *Child Dev.* **62**, 1379–1392.
Kochanska, G. (1993). Toward a synthesis of parental socialization and child temperament in early development of conscience. *Child Dev.* **64**, 325–347.

Kochanska, G. and Murray, K. (2000). Mother–child mutually responsive orientation and conscience develeopment: from toddler to early school age. *Child Dev*. **71**, 417–431.

Lalumiere, M.L. and Quinsey, V.L. (1996). Sexual deviance, antisociality, mating effort, and the use of sexually coercive behaviors. *Personality Individ. Diff*. **21**, 33–48.

Lalumiere, M.L., Harris, G.T. and Rice, M.E. (2001). Psychopathy and developmental instability. *Evol. Hum. Behav*. **22**, 75–92.

Landy, S. and Peters, R. DeV. (1992). Toward an understanding of a developmental paradigm for aggressive conduct problems during the pre-school years. In R. DeV. Peters, R.J. McMahon and V.L. Quinsey (eds), *Aggression and Violence throughout the Lifespan*. Sage, Newbury Park, CA, pp. 1–30.

Lee, C.L. and Bates, J.E. (1985). Mother–child interactions at age two years and perceived difficult temperament. *Child Dev*. **56**, 1314–1325.

Lilienfeld, S.O. (1994). Conceptual problems in the assessment of psychopathy. *Clin. Psychol. Rev*. **14**, 17–38.

Lilienfeld, S.O., Gershon, J., Duke, M., Marino, L., and de Waal, F.B.M. (1999). A preliminary investigation of the construct of psychopathic personality (psychopathy) in chimpanzees (*Pan troglodytes*). *J. Comp. Psychol*. **113**, *365–375*.

Lilienfeld, S.O and Marino, L. (1999). Essentialism revisited: evolutionary theory and the concept of mental disorder. *J. Abnorm. Psychol*. **108**, 400–411.

Lykken, D.T. (1957). A study of anxiety in the sociopathic personality. *J. Abnorm. Soc. Psychol*. **55**, 6–10.

Lykken, D.T. (1995). *The Antisocial Personalities*. Erlbaum, Hillsdale, NJ.

Lykken, D.T. (1998). The case for parental licensure. In T. Millon and E. Simonsen (eds), *Psychopathy: Antisocial, Criminal, and Violent Behaviour*. Guilford, New York.

Lykken, D.T. (2001). Parental licensure. *Am. Psychol*. **56**, 885–894, pp. 122–143.

Lynam, D.R. (1996). Early identification of chronic offenders: who is the fledgling psychopath? *Psychol. Bull*. **120**, 209–234.

Lytton, H. (1990). Child and parent effects in boys' conduct disorder: a reinterpretation. *Dev. Psychol*. **26**, 683–697.

MacMillan, J. and Kofoed, L. (1984). Sociobiology and antisocial personality: an alternative perspective. *J. Nerv. Ment. Dis*. **172**, 701–706.

Magid, K. and McKelvie, C.A. (1987). *High Risk: Children without a Conscience*. Bantam Books, New York.

Massachusetts Department of Correction (2002). MCI-Plymouth (Internet site). Retrieved 27 February 2002 from http://www.state.ma.us/doc/facility/FPlymouth.html

McCord, J. (1983). The psychopath and moral development. In W.S. Laufer and J.M. Day (eds), *Personality Theory, Moral Development and Criminal Behaviour*. Lexington Books, Lanham, MD, pp. 357–372.

McCord, J. (2001). Psychosocial contributions to psychopathy and violence. In A. Raine and J. Sanmartín (eds), *Violence and Psychopathy*. Plenum, New York, pp. 141–169.

McHoskey, J. (1995). Narcissism and Machiavellianism. *Psychol. Rep*. **77**, 755–759.

McHoskey, J. (2001a). Machiavellianism and personality dysfunction. *Personal. Individ. Diff*. **31**, 791–798.

McHosky, J.W. (2001b) Machiavellianism and sexuality: on the moderating role of biological sex. *Personal. Indiv. Diff*. **31**, 779–789.

McHoskey, J.W., Worzel, W. and Szyarto, C. (1998). Machiavellianism and psychopathy. *J. Pers. Soc. Psychol*. **74**, 192–210.

Mealey, L. (1992). Are monkeys nomothetic or idiographic? *Behav. Brain Sci*. **15**, 161–162.

Mealey, L. (1995). The sociobiology of sociopathy: An integrated evolutionary model. *Behav. Brain Sci.* **18**, 523–599.
Mealey, L. (1997). Heritability, theory of mind, and the nature of normality. *Behav. Brain Sci.* **20**, 527–532.
Mealey, L. and Kinner, S. (in press). The perception–action model of empathy and psychopathic 'cold-heartedness'. *Behav. Brain Sci.*
Meehl, P.E. (1992). Factors and taxa, traits and types, differences of degree and differences in kind. *J. Personality* **60**, 117–174.
Meehl, P.E. (1995). Bootstrap taxometrics: solving the classification problem in psychopathology. *Am. Psychol.* **50**, 266–275.
Millon, T., Simonsen, E. and Birket Smith, M. (1998). Historical conceptions of psychopathy in the United States and Europe. In T. Millon and E. Simonsen (eds), *Psychopathy: Antisocial, Criminal, and Violent Behaviour*. Guilford, New York, pp. 3–31.
Moffitt, T.E. (1993). The neuropsychology of conduct disorder. *Dev. Psychopathol.* **5**, 135–151.
Moffitt, T.E., and Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Dev. Psychopathol.* **13**, 355–375.
Newman, J.P. (1998). Psychopathic behaviour: an information processing perspective. In D.J. Cooke, A.E. Forth and R.D. Hare (eds), *Psychopathy: Theory, Research and Implications for Society*. Kluwer, Dordrecht, pp. 81–104.
Newman, J.P. and Wallace, J.F. (1993). Diverse pathways to deficient self-regulation: implications for disinhibitory psychopathology in children. *Clin. Psychol. Rev.* **13**, 699–720.
O'Connor, L.E., Berry, J.W., King, J., Yi, E. and Stiver, D. (2000). A new measure of chimpanzee psychopathology: sex matters. Presented at the Human Behaviour and Evolution Society, Amherst, MA, June.
Ogloff, J.R., Wong, S. and Greenwood, A. (1990). Treating criminal psychopaths in a therapeutic community program. *Behav. Sci. Law* **8**, 181–190.
Patrick, C.J., Bradley, M.M. and Lang, P.I. (1993). Emotion in the criminal psychopath: startle reflex modulation. *J. Abnorm. Psychol.* **102**, 82–92.
Patrick, C.J., Zempolich, K.A. and Levenston, G.K. (1997). Emotionality and violent behaviour in psychopaths: A biosocial analysis. In A. Raine (ed.), *Biosocial Bases of Violence*. Plenum, New York, pp. 145–161.
Patterson, G.R. (1992). Developmental changes in antisocial behaviour. In R. DeV. Peters, R.J. McMahon and V.L. Quinsey (eds), *Aggression and Violence throughout the Lifespan*. Sage, Newbury, CA, pp. 52–82.
Pinel, P. (1806/1962). *A Treatise on Insanity* (M.D.D. Davis, trans.). Hafner, New York.
Preston, S. and de Waal, F. (in press). Empathy: its ultimate and proximate bases. *Behav. Brain Sci.*
Quinsey, V.L., Harris, G.T., Rice, M.E. and Cormier, C.A. (1998). *Violent Offenders: Appraising and Managing Risk*. American Psychological Association, Washington, DC.
Quinsey, V.L., Rice, M.E. and Harris, G.T. (1995). Actuarial prediction of sexual recidivism. *J. Interpers. Violence* **10**, 85–105.
Radke-Yarrow, M. and Zahn-Waxler, C. (1986). The role of familial factors in the development of prosocial behaviour: research findings and questions. In D. Olweus, J. Block and M. Radke-Yarrow (eds), *Development of Antisocial and Prosocial Behaviour: Research, Theories, and Issues*. Academic Press, San Diego, CA, pp. 207–234.

Reid, W.H. (2001). Antisocial personality, psychopathy, and forensic psychiatry. *J. Psychiat. Pract.* **January**, 55–58.

Repacholi, B., Slaughter, V., Pritchard, M., and Gibbs, V. (2001). Theory of mind, Machiavellianism, and peer relationships in childhood. Paper presented at the Macquarie Centre for Cognitive Science Workshop: Mind-reading and Behaviour: Individual Differences in Theory of Mind and Implications for Social Functioning, July 2001, Sydney, Australia.

Rice, M.E. (1997). Violent offender research and implications for the criminal justice system. *Am. Psychol.* **52**, 414–423.

Rice, M.E. and Harris, G.T. (1995). Violent recidivism: assessing predictive validity. *J. Consult. Clin. Psychol.* **63**, 737–748.

Rice, M.E., Harris, G.T. and Cormier, C.A. (1992). An evaluation of a maximum security therapeutic community for psychopaths and other mentally disordered offenders. *Law Hum. Behav.* **16**, 399–412.

Rowe, D.C. (1994). *The Limits of Family Influence: Genes, Experience, and Behaviour*. Guilford, New York.

Rygaard, N.P. (1998). Psychopathic children: indicators of organic dysfunction. In T. Millon and E. Simonsen (eds), *Psychopathy: Antisocial, Criminal, and Violent Behaviour*. Guilford, New York, pp. 247–259.

Sadler, J.Z. (1999). Horsefeathers: a commentary on 'Evolutionary vs. prototype analyses of the concept of disorder'. *J. Abnorm. Psychol.* **108**, 433–437.

Salekin, R.T., Rogers, R. and Sewell, K.W. (1996). A review and meta-analysis of the Psychopathy Checklist and Psychopathy Checklist, Revised: predictive validity of dangerousness. *Clin. Psychol. Sci. Pract.* **3**, 203–215.

Seto, M.C. and Barbaree, H.E. (1999). Psychopathy, treatment behaviour, and sex offender recidivism. *J. Interpers. Violence* **14**, 1235–1248.

Seto, M.C., Khattar, N.A., Lalumiere, M.L. and Quinsey, V.L. (1997). Deception and sexual strategy in psychopathy. *Personal. Individ. Diff.* **22**, 301–307.

Seto, M.C., and Lalumiere, M.L. (2000). Psychopathy and sexual aggression. In C.B. Gacono (ed.), *The Clinical and Forensic Assessment of Psychopathy: a Practitioner's Guide*. Erlbaum, Mahwah, NJ, pp. 333–350.

Shipley, S. and Arrigo, B.A. (2001). The confusion over psychopathy (II): implications for forensic (correctional) practice. *Int. J. Offender Ther. Comp. Criminol.* **45**, 407–420.

Simmons, L.W., Tomkins, J.L. and Hunt, J. (1999). Sperm competition games played by dimorphic male beetles. *Proc. R. Soc. Lond. Ser. B* **266**, 145–150.

Skilling, T.A., Quinsey, V.L. and Craig, W.M. (2001). Evidence of a taxon underlying serious antisocial behaviour in boys. *Crim. Just. Behav.* **28**, 450–470.

Smith, R.J. (1999). Psychopathic behaviour and issues of treatment. *New Ideas Psychol.* **17**, 165–176.

Smith, R.J. and Griffith, J.E. (1978). Psychopathy, the Machiavellian, and anomie. *Psychol. Rep.* **42**, 258.

Snyder, J. and Patterson, G. (1990). Family interaction and delinquent behaviour. In H.C. Quay (ed.), *Handbook of Juvenile Delinquency*. Wiley-Interscience, New York, pp. 216–243.

South Dakota Department of Corrections (2002). Inmate wild fire suppression (media release). Retrieved 27 February 2002 from: http://www.state.sd.us/corrections/CSFireSuppression.htm

Vila, B. (1997). Human nature and crime control: improving the feasibility of nurturant strategies. *Politics Life Sci.* **16**, 3–21.

Wachs, T.D. (1992). *The Nature of Nurture*. Sage, Newbury Park, CA.

Wakefield, J.C. (1992). The concept of mental disorder: on the boundary between biological facts and social values. *Am. Psychol.* **47**, 373–388.
Wakefield, J.C. (1999). Evolutionary prototype analyses of the concept of disorder. *J. Abnorm. Psychol.* **108**, 374–399.
Wakefield, J.C. (2000). Aristotle as sociobiologist: the 'Function of a Human Being', black box essentialism, and the concept of mental disorder. *Phil. Psychiat. Psychol.* **7**, 17–44.
Weiss, A., King, J.E. and Figueredo, A.J. (2000). The heritability of personality factors in chimpanzees (*Pan troglodytes*). *Behav. Genet.* **30**, *213–221*.
Widiger, T.A. (1998). Psychopathy and normal personality. In D.J. Cooke, A.E. Forth and R.D. Hare (eds), *Psychopathy: Theory, Research and Implications for Society*. Kluwer, Dordrecht, pp. 47–68.
Widiger, T.A., Cadoret, R., Hare, R., Robins, L., *et al.* (1996). DSM-IV antisocial personality disorder field trial. *J. Abnorm. Psychol.* **105**, 3–16.
Widiger, T.A. and Clark, L.A. (2000). Toward DSM-V and the classification of psychopathology. *Psychol. Bull.* **126**, 946–963.
Widiger, T.A., and Lynam, D.R. (1998). Psychopathy and the five-factor model of personality. In T. Millon and E. Simonsen (eds), *Psychopathy: Antisocial, Criminal, and Violent Behaviour*. Guilford, New York, pp. 171–187.
Widiger, T.A. and Sankis, L.M. (2000). Adult psychopathology: issues and controversies. *Ann. Rev. Psychol.* **51**, 377–404.
Williamson, S.E., Harpur, T.J. and Hare, R.D. (1991). Abnormal processing of affective words by psychopaths. *Psychophysiology* **28**, 260–273.
Wilson, D.S. (1998). Adaptive individual differences within single populations. *Phil. Trans. R. Soc. Ser. B* **353**, 199–205.
Wong, S. and Hare, R.D. (in press). *Program Guidelines for the Institutional Treatment of Violent Psychopaths*. Multi-Health Systems, Toronto.
Zachar, P. (2000). Psychiatric disorders are not natural kinds. *Phil. Psychiat. Psychol.* **7**, 167–182.
Zuckerman, M., Buchsbaum, M.S. and Murphy, D.L. (1980). Sensation seeking and its biological correlates. *Psychol. Bull.* **88**, 187–214.

17

Borderline Personality Disorder and Theory of Mind: an Evolutionary Perspective

GERHARD DAMMANN

Department of Psychiatry and Psychotherapy, University of Basle, Switzerland

The history of the development of a theory of mind is tightly bound to the development of a social environment of evolutionary adaptedness (EEA). According to most evolutionary psychologists, human psychological adaptations can be recognised by criteria such as high efficiency, high complexity, high modularity, low phenotypic variance, low genotypic variance, low hereditability, universality across cultures, and universality across individuals. Adaptationist criteria, even for the development of a theory of mind, must recognise two typical kinds of psychological adaptations: naturally selected survival mechanisms and sexually selected fitness indicators (Miller 2000). The mentalisation capacity appears to originate from the basis of attachment, which also represents an important middle-level theory. The term 'theory of mind' was coined by Premack and Woodruff (1978) and refers to the individual's capacity to make inferences about the mental states in oneself and in others, in order to understand and predict behaviour.

The significance of the acquisition and development of a theory of mind for the young child has been uncontested, in line with theorising on hominid evolution (Moore and Frye 1991). Thus, the importance in untangling the onset and the particular antecedents that lead to a theory of mind is critical. Moreover, such an endeavour throws light not only on the particulars of normal development, but also has great clinical implications for understanding

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

atypical development in the case of autism (Baron-Cohen, Leslie and Frith 1985) or other forms of psychopathology, such as the borderline personality (Fonagy 1989, 1991).

BORDERLINE PERSONALITY DISORDER

Borderline personality disorder (BPD) is a severe, chronic, disabling and potentially lethal psychiatric condition. People who suffer from this disorder have extreme and long-standing instability in their emotional lives, as well as in their behaviour and their self-image. This is a common disorder affecting 2% of the general population. The best evidence indicates that about 11% of psychiatric outpatients and 19% of inpatients meet diagnostic criteria for BPD (Kass *et al.* 1985).

These instabilities of emotion, behaviour and self-image have devastating and sometimes deadly consequences. People with BPD have repeated and frequent difficulties in their relationships and work lives and they feel alternating extremes of anger, depression and emptiness. All too frequently, 69–75% of individuals with BPD resort to self-destructive behaviours, such as self-mutilation, alcohol and drug abuse, serious over- or under-eating and suicide attempts, to try to escape from their emotional turmoil (Clarkin *et al.* 1983; Cowdry, Pickar and Davies 1985). The completed suicide rate for BPD individuals is 3–9.5% (McGlashan 1986; Stone 1983), which is comparable to the other serious psychiatric disorders, such as depression, alcohol dependence and schizophrenia.

The seriousness of BPD is compounded by the fact that it is difficult to treat. The very characteristics of the disorder, such as unstable relationships and intense anger, interfere with establishing the therapeutic relationship that is necessary to any treatment, whether psychotherapy or medication. Further, mental health professionals often are reluctant to treat these individuals because they exhibit two characteristics likely to lead to clinician 'burnout': BPD persons' hostility towards the clinical professional, and their persistent suicidal thoughts and feelings (Hellman, Morrison and Abramowitz 1986).

Despite the devastating nature of this disorder, it has not received the scientific and clinical attention that other health and psychiatric problems of equal, or even lesser, degrees of disability have received.

There is a lot of evidence that BPD occurs significantly more often in a socially detached or disorganised environment: lower-class environments, family violence, alcohol problems among parents, early separation experiences, high number of siblings, stepfathers or stepmothers, sexual abuse. Among preschool children, child abuse is 40 times greater in stepfamilies than in families with two genetic parents (Daly and Wilson 1987).

MENTALISATION AND ATTACHMENT

Extensive contact between infants and caregivers (especial mothers) immediately after birth may promote emotional bonding. Mothers make great efforts to establish eye contact with their infants. When infants reciprocate eye contact, mothers become livelier, speak with greater voice inflections and approach the infants more closely (Klaus and Kennell 1976).

'The child's early social environment, mediated by its primary caregiver, directly influences the ontogenetic development of the structure of the brain, on which the child's further social development depends' (Schore 1994, p. 62). In the centre of Schore's model stands the region of the ventromesial frontal lobe. In accordance with his concept of the inhibiting component of a comprehensive motivation system, this demonstrates that its origin comes from the subcortical structures of the limbic region and of the brainstem and its neuroanatomical maturation, which, as he shows, is considerably shaped by certain aspects of the early mother–child interaction. In this respect, one could almost argue that, with regard to its mentalisation function, the internalised mother is embodied in the ventromesial frontal lobe.

Mentalisation theory is tightly bound to other concepts, especially to memory theory, affect-regulation theory, the development of the self, consideration of (language) narrative processes, as well as the development of recognition and interpretation of expressions and affective behaviour. A large role is attached to attentiveness and decision-making processes (see e.g. the somatic marker hypothesis, proposed by Damasio 1995). In clinical situations, BPD patients can, on the one hand, sometimes manifest impulsive behaviours, as if they are always 'in a rush', while, on the other hand, they are more often inhibited in making real decisions, due to the typical ambivalences in personality disorders.

Following Wright (1994), attachment theory has two components: a normative component, which explains modal, species-typical patterns and stages of attachment, e.g. higher theory of mind in humans than in non-human primates; and an individual-difference component, which accounts for deviations from the modal patterns and stages. 'This bifurcation — explaining the species-typical patterns of attachment behaviour and individual differences that diverge form these patterns — is a hallmark of modern evolutionary models' (Simpson 1999, p. 122).

Several authors have implied that the mentalisation ability and the related complex of social group behaviour strategies are responsible for neoteny (the prolongation of the juvenile non-reproductive period among human beings) (Gordon 1989; Joffe 1997). The contained mentalisation ability of the mother always also represents a model for social behaviour strategies, which the child mentalises and takes on, not only as socially learned behaviour.

From an evolutionary psychological perspective, too much cognitive function can also prove to be a handicap (see Schmitt and Grammer 1997). Worth noting in this context is the theoretical consideration proposed by Brüne (2001) concerning whether the brain's extreme adaptation to social conditions might not also be an overspecialisation, which could explain in part the appearance of certain psychopathological processes as human characteristics.

Daniel Stern (1985) raises the question of how the repeated, invariant features of experience that have been identified — the experiences of agency, coherence, affectivity and memory — become integrated into the organising perspectives that are characterised as the sense of self and of others. Stern proposes that memory for repeated episodes provides the basis for such integration. Protypic memory structures ('the generalised breast-milk episode') involving actions, sensations and affects that occur in a temporal, physical and causal relationship and in an interactive, interpersonal context form the basis of the non-verbal representational schemas that Sterns terms 'representations of interactions that have been generalized (RIGs)'. According to Lewis and Brooks (1975) or Beebe and Lachmann (1988), the 'infant represents the distinctive features of social interactions before they are abstracted and before they are symbolised' (Beebe and Lachmann 1988, pp. 310–311). 'The capacities necessary to develop these structures of cognition and social interaction are in place in the first year of life, long before language is acquired' (Bucci 1997, p. 150).

The Anglo-Hungarian psychoanalyst and developmentary psychopathologist, Peter Fonagy, has offered an approach (Fonagy *et al*. 2002, p. 126) whose core proposition is the rejection of the notion that the conscious apprehension of our mind states through introspection might be a basic, direct and probably prewired ability of our mind. Fonagy *et al*. do not believe that knowledge of the self as a mental agent is innately given. Rather, they see it as a developing or constructed capacity that evolves out of the earliest relationships. Their core idea is that the attachment context provides the setting in which the infant can develop a sensitivity to self-states, through what Gergely (1994) has termed 'psycho-feedback' or social biofeedback.

Today, 'life history theory' (e.g. Stearns 1992) has emerged as a major perspective in evolutionary thinking. To leave descendants, individuals must solve problems of survival, growth, development and reproduction across the lifespan. Attachment (and mentalisation) are important for both sides of reproductive effort: mating and parenting. In general, evolutionary theories (reciprocal altruism, attachment, sexual selection, etc.) deal with the different problems of adaptation and of social behaviour.

Bowlby's original concept has been elaborated by Inge Bretherton, P. Crittenden, Mary Main (1981), L. Sroufe and others. Main has conjectured that different patterns of attachment in children might represent different evolutionary strategies for enhancing inclusive fitness in certain environments.

Four representational systems are implied in these reformulations of Bowlby's evolutionary attachment theory (see Fonagy *et al*. 2002, p. 123):

1. Expectations of interactive attributes of early caregivers created in the first year of life and subsequently elaborated.
2. Event representations by which general and specific memories of attachment-related experiences are encoded and retrieved.
3. Autobiographical memories by which specific events are conceptually connected because of their relation to a continuing personal narrative and developing self-understanding.
4. Understanding of the psychological characteristics of other people (inferring and attributing causal motivational mind states, such as desires and emotions, and epistemic mind states, such as intentions and beliefs) and differentiating these from those of the self.

Bowlby and Ainsworth originally believed that the secure pattern of attachment was 'nature's prototype'. 'Security may have been the most common pattern of attachment in evolutionary history. However, selection pressures should not have generated a single prototype' (Simpson 1999, p. 126). 'There is no best mothering (or attachment) style, for different styles are better in different circumstances . . . optimal mothering (and attachment) behaviour will differ according to the sex of the infant, its ordinal position in family, the mother's social status, caregiving contributions from other family members, the state of physical resources and so on' (Hinde 1982, p. 71). Nonetheless, security may be the primary or default survival strategy of the attachment system if environmental conditions are suitable (Main 1990). The attachment style constitutes itself, in any case, at least as much through the value which is generally attached to relationship experiences as through the relationship experience itself.

A number of long-term studies have shown that a secure attachment in early childhood influences strongly and favourably many aspects of adaptation, e.g. social behaviour (Skolnick 1986), affective regulation (Erickson, Scroufe and England 1985) and cognitive talent (Matas, Arend and Scroufe 1978; Grossmann and Grossmann 1991).

INTERNAL WORKING MODELS AND THEORY OF MIND

One theoretical basis of attachment relationships is the construct of Bowlby's internal working models (or multiply-connected schema networks).

> Beyond infancy, attachment relations come to be additionally governed by internal (or mental) working models that young individuals construct form the experienced interaction patterns with their principal attachment figures. These internal working models are conceived as 'operable' models of self

> and attachment partner, based on their joint relationship history (Bretherton and Munholland 1999, p. 89).

Analogous to object relation theory, internal working models can help in the reconstruction of how expectations based on the interactions with the primary caregiver arise in repeatedly experienced interaction patterns. According to Daniel Stern (1994), the repeated childhood experience of being picked up after falling down, which becomes encoded as 'a-schema-of-being-with' with the mother, leads to the expectation that reassurance and consolation will follow distress or fear (Fonagy 1996).

Bucci (1997) describes the converging data on the development of symbolising function in the infant's cognitive and emotional world:

> Mandler's [1992] concept of perceptual analysis, leading to the development of image-schemas, parallels Kosslyn's (1987) analysis of the means by which perceptual information in continuous processing systems is chunked or funnelled through functionally equivalent classes of representation to prototypic imagery, including representations of relations as well as objects. These concepts converge also with Stern's (1985) formulation of emotional organisation through development of abstract, prototypic experiential schemas (p. 151).

The evolutionists Belsky (1999) or Shaver, Collins and Clark (1996) discuss two perspectives on multiple working models. The traditional perspective is that children develop a hierarchy of working models and that the primary one exerts the most influence.

> Instead, though, it may be the case that alternative models provide the developing individual with multiple reproductive templates that can later be enacted, depending upon the mating and parenting (i.e. reproductive) circumstances he or she subsequently encounters. Thus, patterns of attachment that do not develop early in life may simply not be available for use as reproductive strategies (Belsky 1999, p. 157).

The two models are not exclusive. A hierarchy may exist, Belsky argues, with the primary model serving as the default option and the secondary model coming 'on-line' only when the primary model fails.

Overall, a multitude of theories have been proposed as explanatory frameworks for the emergence of the 'theory of mind'. These theories can be distinguished between those that credit the person with an acquisition of a 'theory', and those that assume that an understanding of the contents of the mind is dependent on the ability to simulate others (e.g. Gordon 1986).

On a different level, the theories of 'theory of mind' can be classified among those that propose an acquisition of a theory of mind as a result of an innately pre-programmed maturation of domain-specific modules (e.g. Leslie 1987; Baron-Cohen 1995), and those that support an experience-based acquisition (Hobson 1993).

THE SOCIAL EEA

Brewer and Caporael (1990) have proposed that participation in the daily functioning of small cooperative groups may have been a principal survival strategy of early humans. The human mind was designed by natural selection to deal with the major adaptive problems humans faced in evolutionary history, especially the social EEA (Tooby and Cosmides 1992).

> Psychological mechanisms are believed to be evolved solutions to adaptive problems if they show evidence of 'special design' — that is, if they provide a precise, specialised, efficient, economical and reliable solution to a specific adaptive problem. Some psychological mechanisms presumably evolved in response to stable features of social EEAs, such as the recurrent structures, rules, roles and perils associated with living in small groups. Though small cooperative groups should have facilitated the inclusive fitness of most group members by providing greater protection form predators and better access to mates and food, group living should have posed some unique adaptive problems. To minimise the adverse effects of 'cheaters', for example, humans have apparently developed a keen ability to detect people who do not reciprocate equitably in groups over time (Simpson 1999, p. 121).

Of course the social EEA was 'neither as uniform nor as benign as Bowlby seems to have imagined' (Chisholm 1996, 14):

> Once it is acknowledged that there were many different EEAs and it is accepted that attachment behaviour probably evolved in response to varying selection pressures over the course of human history, it becomes increasingly difficult if not untenable to embrace the notion that one pattern of attachment (i.e. proximity-seeking and contact-maintaining security) was or is 'species-typical' (Belsky 1999, p. 143).

The development of the adapted mind is connected with the development of the social EEA. For thousands of generations men probably lived, as hunters and gatherers, in small cooperative groups.

> Most people in a tribe were biologically related to one another, and strangers were encountered rather infrequently, probably during periods of intertribal trading or war (Simpson 1999, p. 121).

> Infants do not have the cognitive ability to appraise the 'quality' of local environmental conditions (e.g. whether the local environment is safe, plentiful and rich in resources vs. threatening, harsh and impoverished). However, they do have the ability to discern whether their caregivers are providing them with the level of sensitivity, responsiveness, and attention dictated by their biological needs (Simpson 1999, p. 123).

The child thus 'relies on' the ability of the mother to recognise and avoid dangers on behalf of the child (one possibly finds among mothers whose

children are involved in accidents, e.g. drowned in ponds, increased numbers of people with borderline disorders).

When talking to their infants, mothers slow their speech, accentuate certain syllables and often talk one octave above normal speech. This pattern is known as 'motherese' (Grieser and Kuhl 1988).

Following Simpson (1999), Mary Ainsworth's famous 'strange situation paradigm' is well suited to detect different patterns of attachment because it presents infants with two common 'cues of danger' in the social EEA: being left alone and being left with a stranger. Each attachment pattern reflects a different ecologically contingent strategy designed to solve adaptive problems posed by different rearing environments (of course, children with secure attachment prefer being with the mother — the safest situation — but are also able to be alone or with a stranger). Mothers of securely attached infants are available and responsive to the needs and signals of their infants.

ATTACHMENT AS A DETERMINANT OF CAPACITIES FOR SOCIAL COGNITION

Peter Fonagy, whose theory I will critically engage below, presented in January 2000, at the Congress of the International Association of Adolescent Psychiatry in San Francisco, the perspective of the attachment theorist as follows:

> Attachment theory postulates that early experiences with the parent provide prototypes for all later relationships mediated by so called 'internal working models' (Bowlby 1973, 1980). The Adult Attachment Interview (AAI; George, Kaplan and Main 1985) was designed to provide a classification of these, analogous to the Strange Situation classification. The instrument elicits narrative histories of childhood attachment relationships. The AAI scoring system (Main and Goldwyn 1994) classifies individuals into Secure/Autonomous, Insecure/Dismissing, Insecure/Preoccupied or Unresolved with respect to loss or trauma, categories based on the structural qualities of narratives of early experiences. While autonomous individuals value attachment relationships, coherently integrate memories into a meaningful narrative and regard these as formative, insecure individuals are poor at integrating memories of experience with the meaning of that experience. Those dismissing of attachment show avoidance by denying memories, or by idealising or devaluing early relationships. Preoccupied individuals tend to be confused or angry in relation to attachment figures, often still complaining of childhood slights, echoing the protests of the resistant infant. Unresolved individuals give indications of significant disorganisation in their attachment relationship representation, through semantic or syntactic confusions in their descriptions of childhood trauma or a recent loss. Secure adults are three or four times more likely to have children who are securely attached to them (van Ijzendoorn 1995). This is

true even where parental attachment is assessed before the birth of the child (Steele, Steele and Fonagy 1996; Ward and Carlson 1995).

Attachment theorists have assumed that securely attached adults are more sensitive to their children's needs, thus fostering an expectation in the infant that dysregulation will be rapidly and effectively met (Belsky, Rosenberger and Crnic 1995; De Wolff and van Ijzendoorn 1997). Disappointingly, standard measures of caregiver sensitivity do not appear to explain at all well transgenerational consistencies in attachment classification (van Ijzendoorn 1995). An alternative view is provided by Mary Main (1991) and Inge Bretherton (1991), who independently drew attention to what the philosopher Dennett called the 'intentional stance'. Dennett (1987) stressed that human beings try to understand each other in terms of mental states, that is, thoughts and feelings, beliefs and desires, in order to make sense of and, even more important, to anticipate each others' actions. If the child is able to attribute an unresponsive mother's apparently rejecting behaviour to her sadness about a loss, rather than simply feeling helpless in the face of it, the child is protected from confusion and a negative view of himself. The hallmark of the intentional stance is the child's recognition at around 3–4 years that behaviour may be based on a mistaken belief (Fonagy 2000).

REFLECTIVE FUNCTIONING

From an evolutionary perspective, the first question arises as to what advantages are connected with a 'surplus' of intelligence, such as one finds among higher primates and human beings. Following the pure cognitive capability, the mentalisation capacity of human beings and higher primates comes quickly into focus (Hare, Call and Tomasello 2001). Among others, Brothers (1990) and Dunbar (1998) developed the 'hypothesis of the social brain'. Until now, however, almost no connection between the 'social brain hypothesis' and evolutionary attachment theory was able to be made (Simpson 1999; Belsky 1999).

'According to the current state of knowledge, human beings' social meta-cognitive capacities developed over gradual stages in the course of primate evolution' (Brüne and Ribbert 2001, p. 57). With respect to reflective functioning, it is a matter of a construct that is illuminated from psycho-analytical, evolutionary psychological (Lorenz 1973) and cognitive scientific perspectives. It applies to the ability, developed in the course of evolution, of a person to discern and to understand in himself, as well as in others, concepts of intentionality (internal intentions) and mental states (feelings, thoughts, desires, etc.), as well as to reflect appropriately on the witnessed behaviour. For Fonagy, 'mentalisation' or 'reflective function' shows the capacity to understand personal and foreign behaviours in the form of 'mental states' (feelings, desires, etc.). For example, a child at the age of 3–4 years recognises

that another person's behaviour can be based on an erroneous assumption. This fact may be seen through a simple example: if one shows a 3–4 year-old child a package of candy, he will say that there is candy in the package. If one shows him that there is in fact a pencil in the package and that he erred and then asks him what his playmate who is waiting outside will answer to the same question, the child — if he has already attained the capability to perform this kind of mentalisation — will answer 'candy' and thus be able to anticipate the probable error of his pal. He will, however, answer 'pencil', what he himself experienced, if he is still in an earlier phase of reflective functioning. This concept also naturally corresponds to Jean Piaget's model of the change of perspective. Recent findings, mainly from observational studies (Dunn 1988) and from novel experimental designs, throw light on 'the dark age' preceding the child's fourth birthday. The latter allows 3 year-olds to perform at a better level, and hence display an understanding on the distinction between true and false beliefs (see Moses and Chandler 1992). The underlying aim of such studies is to surpass the assumed cognitive deficits of the children, which hinder their ability to pass the false-belief tasks, in the quest of uncovering the antecedents of theory of mind.

In the literature there is a host of concepts which partially overlap each other: metacognition, metacognitive control, mentalisation, theory of mind, reflective self function, enphronesis, capacity for symbolisation and perspective change. The new area of developmental psychopathology, which has been shaped in the last few years by Dante Cicchetti, Robert Emde, Michael Rutter, Mary Main and Peter Fonagy, and for which the attachment theory based on evolutionary medicine constitutes a central foundation, offers the possibility to mediate between individual psychopathology and individual and anthropological development.

Following clinical 'theory of mind' and attachment theory, the thinking of the child is formed or destroyed in the process of its creation through the thinking of the primary caregiver (Fonagy 1991). Secure attachment is the basis of the acquisition of metacognitive or mentalisation capacity; the caregiver's capacity to mentalise may foster the child's bonding with the parent; abuse or neglect may undermine the acquisition of a mentalisation capacity; symptoms of BPD may arise as a consequence of inhibited mentalisation. Not only is attachment, therefore, a fundamental need but so also is the attempt of people in their 'mental states' (i.e. feelings, thoughts, desires, etc.) to understand each other. This fundamental function ('reflective functioning') prevents confusion and difficulties related to self-worth. Continuity thereby contributes to the building of mental structures.

> In summary and consistent with the evolutionary perspective under consideration, there are grounds for tentatively concluding that secure attachment in childhood may be a central part of a developing, facultative

> reproductive strategy designed to promote a quality-vs.-quantity orientation toward reproduction (Belsky 1999, p. 153).

Since 1996, our research group in Munich has studied the predictive value of evolutionary psychologically and human ethologically-orientated concepts for the appearance of post-partum affective disorders (post-partum depression and 'baby blues'). We have been able to explain that precisely both adaptive or functional aspects (stronger in the case of dysphoria) and maladaptive or dysfunctional aspects (stronger in the case of depression) play a role in such disorders (Dammann *et al.* 2001; Dammann and Schiefenhövel in preparation). Post-partum depression actually appears in many respects nearly to demonstrate the concepts of attachment and mentalisation discussed here, since it is known that a mother's post-partum depression can lead to problems in the cognitive development of the child (e.g. as seen in the survey by Murray and Cooper 1997).

Through the inclusion of the processes of metacognitive control, the research group could further explain the transition gap of Fonagy's attachment theory. Fonagy (1997) points out that most studies confound two independent psychological processes, which each present their own determinants of juvenile attachment ability. The first process level refers to the mother's attitude and representation of attachment and to her behaviour (sensitivity), in other words variables that exist independently of the child's psychological condition. The second process level refers to the ability of the mother to imagine her child as a cognitive being, as a person with intentions, feelings and desires. This requirement demands that she reflects on the psychological health of another human being and thereby that she goes beyond simple attention and affection [metacognition or capacity for (self-)reflectiveness, following Daudert 2001, p. 50].

REFLECTIVE FUNCTIONING AND PERSONALITY DISORDERS

Personality disorders include deeply rooted, continuous patterns of behaviour, which appear in inflexible reactions to various life situations without evidencing the existence of psychosis. In the foreground usually stands a considerable disorder of relationships, but also disorders of the concept of self or of identity, which are described as minimally flexible. Patients with personality disorders either suffer themselves in their personality or provoke suffering in their environment (as with antisocial personality disorders).

Unfortunately, adaptive and maladaptive aspects of disturbances in the mentalising capacity of persons with severe personality disorders have been until now infrequently discussed from an evolutionary psychological and

human ethological view. Therefore, attachment theory as applied to BPD needs first to be presented in greater detail, the limits to this approach discussed, alternative models critically considered, and possible evolutionary psychological conclusions as applied to the development of defects in mentalisation in cases of severe personality disorders evaluated.

As with many BPD patients, persons whose childhood was shaped by extremely threatening or neglectful surroundings can realise a 'benefit', in the sense of an early defence mechanism, from the renunciation of the 'structure-creating values' gleaned from identification with the 'mental states' of the other. The term 'theory of mind' covers not only the capacity to think about one's own thoughts, but also mental abilities that have been variously labelled 'mentalisation', 'reflective-self capacity', 'metacognition' or 'seeing from another's perspective'. According to this theory, early, traumatising experiences lead, therefore, to an inhibition in the mentalisation ability, which later creates countless difficulties.

This defence mechanism of 'mentalisation' can occur, for instance, through dissociated conditions. In the 1990s, Liotti drew attention to the connection between disorganised attachment types and 'dissociation'. Liotti (1992) suspected that fear-creating or fear-fulfilling behaviour by the person to whom the patient is attached leads to a paradox. The paradox was that the person needed to flee to and flee from the object of their attention, which also corresponded to unresolved attachment representation. The dissociation presents a 'solution' to the conflict, in so far as the person quasi-opts out of the contradictory demands. To this conclusion also fit the results that a person with an unresolved, disorganised attachment style exhibits an increased ability to be hypnotised and that suggestibility and hypnotisability are correlated with traumatisation (Solomon and George 1999).

The close connection between violence, abuse and neglect and deficits in reflective functioning has been proved in countless studies (Cicchetti and Lynch 1995). As an expression of the lacking ability to change perspective and as a result a certain form of 'lack of imagination', the contrast between heightened sensitivity of the BPD patients (i.e. an increased receptivity) and their lacking sensibility (i.e. true empathy for the world of the other person) has been clinically impressive. The mental functioning of patients suffering from BPD is characterised by unintegrated representations of self-with-other, emotional dysregulation and serious deficits in self-reflective and metacognitive capacities. For reasons of self-coherence, BPD patients are left only with the possibility to 'externalise' experienced but unintegrated representations. Successful mentalisation makes it possible to distance oneself or to remain alone. Exactly because psychological closeness has become so unbearable, BPD patients will continue to seek out the bodily closeness of an abuser. Manifold clinical symptoms (especially diffusion of identity and emotional and interpersonal instability) can be understood with this theory (Dammann 2001).

Psychotherapeutic work with these patients may facilitate the reactivation of this inhibited capacity. The therapist's mentalistic, elaborative 'intentional stance' (Dennett 1987) enables the patient to find him/herself in the therapist's mind as a thinking, feeling being.

All personality disorders are characterised by persons being often exceptionally sensitive (e.g. in the case of sickness) and by an increased difficulty, despite this sensitivity, in realistically estimating the motivations or intentions of the other person. Hence, these factors create difficulty in taking on the perspective of the other person. With the help of a theory of mind, a child can understand the feelings of another person without having to take on those feelings. The ability to tolerate the state of one's own feelings, as with self–object differentiation, is a result of this process and a basis for affect regulation, something which is typically disturbed in BPD patients.

Generally in the literature, the value of mentalisation and its connected abilities are merely emphasised, as well as the disadvantages that occur from a deficit in mentalisation. It is stressed in the psychoanalytical literature that a secure attachment experience promotes the ability to symbolise, which in turn presents the basis for adoption of a theory of mind. Hence, the possibility arises of the experience of an intersubjective understanding of common feelings, which should be the pre-condition for satisfactory communication. The literature has nevertheless failed to appreciate (see below) that first, this ability is only valuable with an appropriate other person, and second, it is connected to conflicts arising from misunderstandings, which grow whenever more aspects become perceptible and thinkable. In my opinion, the absence of mentalisation also offers considerable protective value, especially when a parental figure who does not have or offer this ability is involved.

The following presents two examples of the advantages and disadvantages of different reflective function (RF) qualities (example taken from Daudert 2001, pp. 133–134). First, I present an example of the lacking RF applied to the question 'Since your childhood, have you witnessed any changes in your relationship with your parents?'

> Earlier it was difficult with my father. He criticised me often. But, these were not easy times. Now, we have a good relationship. We see each other sometimes during the holidays and have nice meals together. Between us it has become pretty relaxed.

This example shows a concrete, generalised explanation of behaviours that is not applied to psychological health. Nevertheless, one can imagine that this form of representation has some value, because it functions to create some distance. The following is an example of a existing, average RF applied to the same question:

> I believe that my father has in the meantime understood that my bulimia was there to provoke attention and recognition. I wanted at the time to test

> whether he really loved me. As far as that is concerned, I am now no longer under so much pressure.

This example shows an explicit reference to the psychological conditions. In comparison to the first example, it is also clear that this perspective constitutes an improvement, but also can contain significantly more (neurotic) potential for conflict. Such an explicit reference appears 'sensible' only if the other person is, or was, not too traumatised (by 'sensible' is meant, as always in evolutionary psychiatry, a favourable or adaptive functional strategy).

Clarkin and Kernberg (2002) studied a group of nearly 1000 persons from which 45 demonstrated temperamentally or phenomenologically an emotionally instable and impulsive pattern, which is typical of BPD. Only nine of these 45 patients also fulfilled the criteria of a borderline disorder with severe relationship or behavioural disorders. Those out of the 45 who had really demonstrated or developed a borderline disorder showed simultaneously a diffusion of identity (in addition to distinctly temperamental features). It can therefore be carefully concluded that a high degree of impulsiveness does not alone lead to the development of a personality disorder; rather, that problems with the integration of self-representation have a decisive ability to contribute to their occurrence.

Fonagy *et al.* (1996) were able to show in the Cassel Hospital Study that severely traumatised patients only develop BPD when simultaneously their reflective abilities are only minimally present. Out of 24 substantially traumatised persons who, however, had high RF values, only four showed BPD.

In a recent study, Daudert (2001) was unable to confirm the expected hypothesis that patients with traumatic childhood experiences (such as deprivations), but who nevertheless possessed a share of secure attachment, demonstrate a higher degree of self-reflectivity than traumatised patients without a secure attachment style. Like other studies, this study relativises a perspective that tries to equate the higher mentalisation or reflective functioning with a secure attachment style or secure attachment representation. Main (1991) was able to prove empirically that there are important connections between the quality of the mother's metacognitive monitoring and the development of the child's attachment structures. Interestingly, Daudert's study (2001) was found to contain not only a large group (37.5%) of BPD patients with a deficit in reflective functioning, but also a large group with an impairment of reflective functioning typical for BPD, namely the category of 'over-analysing, hyperactive RF'.

The significance of reflective functioning, mentalisation, or, to be precise, identity diffusion (if one understands it in Otto Kernberg's sense of a lack of an integrated concept of self) for the development of BPD has been

impressively proved in several studies. In the meantime, countless psychodynamic theories and studies (especially Fonagy, Emde, Diamond), as well as cognitive-behavioural ones (above all Liotti), have been able to prove a tight connection between traumatising childhood experiences, social neglect, and disorganised attachment style, especially in the case of a disturbed self-image, in other words BPD. One could therefore describe BPD as a paradigmatic model of a heavy disorder of mentalisation ability, identity creation and affect regulation, which arises overwhelmingly through life circumstances (and not due to genetic factors, which have at most a modulating effect) and does not represent a psychosis or an organic brain disorder.

Traumatised and disorganised-attached children and adults with reduced megacognitive abilities stand especially at risk of developing personality disorders (especially BPD) because of their unstable self-image and their reduced ability at symbolisation. Out of the needs created by the internal confusion and chaos caused by traumatic experiences or neglect, needs which they are not able to reflect upon but which they can understand and even integrate as contradictory self-conditions, such people frequently create new damaging relationships (repeated compulsion), with further fatal results on their internal working model and their future ability to have relationships (Daudert 2001, p. 71). The disorder thus perpetuates itself. Borderline patients are also left alone with their representations of fear or of being persecuted. Moreover, the limited mentalisation ability leads to a lack of consolation, the inability to be alone, along with an experience of deep isolation or loneliness. For, according to the psychodynamic theoreticians of mentalisation (Fonagy and Target 1995), in order for someone to be able to feel truly connected, he/she must first be represented as a psychological being.

According to the so-called 'theory of mind' and attachment theory, a child's thinking is formed through the thinking of the primary caregiver. The latter, especially the mother, names and contains (W.R. Bion 1962) the child's early feelings (which would otherwise be unendurable). Thinking (which may also be understood as affective modulation) is therefore always also the thinking of the other, who appears always represented as an object. If the primary caregiver is not prepared to put herself in the position of the child (because she herself is too fearful, too self-absorbed, or too traumatised, etc.) in order to answer its demands and to make mentalisation ability available, this development will remain rudimentary. The patient cannot contain (unsymbolised) feelings, i.e. cannot console herself. The other object will thereby not be fully integrated, but instead remain as a foreign, threatening lacuna (in the theory's language, as 'the alien other'). The task of the therapist is therefore to make up for this internalised step as the patient shares what is going on in her mind.

CHRONIC BURDEN

This is not the point to go into the neurobiological changes from chronic stress or traumatisation (see e.g. the work of Bessel van der Kolk 1994). Research in this area suggests how changes in the functioning of the amygdala, in the tendency to dissociation, in memory, and in the ability to sense pain properly, create a basis in the central nervous system to overcome traumatisations in the short term, and often impairment in the long term. It is assured that representations of attachment also have an effect on biological processes. Spangler and Grossman (1993) found that when an increase in cortisol was induced through a separation experiment, there was a significantly delayed retrogression of the cortisol levels in insecure persons, and especially in those with disorganised attachment. Gary Kraemer (1992) collected a wealth of results in his research on the psychobiology of attachment among non-human primates. Apes with a so-called 'isolation syndrome' were found to have the following behaviour patterns: self-injury behaviours, hyperphagia and polydipsia, sexual disorders (such as compulsive masturbation), impulsiveness and extreme irritability, delaying during habituation experiments, and much more. Such symptoms can with little effort be imagined among BPD patients.

EVOLUTIONARY PSYCHOLOGY OF ATTACHMENT AND MENTALISATION

> It has been assumed that animals were in the first place rendered social, and that they feel as a consequence uncomfortable when separated from each other, and comfortable whilst together; but it is a more probable view that these sensations were first developed in order that those animals which would profit by living in society, should be induced to live together . . . for with those animals which were benefited by living in close association, the individuals which took the greatest pleasure in society would best escape various dangers; whilst those that cared least for their comrades and lived solitary would perish in greater numbers (Darwin 1871/1981, Vol. 1, p. 80).

As Jeffrey Simpson (1999, p. 115) noticed, Darwin may have been the first attachment theorist, in some respects, as seen in this quotation, although he focused on 'society' (instead of significant others) and 'comrades' (instead of attachment figures). As Simpson (1999) discussed, individual differences in attachment during adolescence and adulthood may reflect different reproductive strategies designed by evolution to enhance reproductive fitness in certain environments. I would like to argue that exactly the same is true for the degree of reflective functioning, or mentalisation. Patterns of different degrees of mentalisation represent nascent facultative reproductive strategies that evolved to promote reproductive fitness in particular ecological niches.

> The concept of 'inclusive fitness' places attachment theory at the center stage of evolutionary sociobiology as a key behavioral mechanism mediating the establishment of genetic proximity, for attachment is the process that ensures that we know whose survival will advantage the reproduction of our genes (Fonagy *et al.* 2002, 122).

This suggests that mentalisation ability can have an influence on all the important problems related to inclusive fitness — survival to reproductive age, mating and reproduction, and raising offspring. Also fitting is the finding of several studies that BPD patients frequently remain without partners or offspring, but nevertheless frequently work as teachers or caregivers. Kunce and Shaver (1994) found that resistant women reported the highest levels of compulsive caregiving.

Trivers (1974) conjectured that the intensity and duration of parent–offspring conflict should depend on factors that affect the cost:benefit ratio over time. Conflict should be heightened when half-siblings exist in families. Because half-siblings share only 25% of their genes, four half-siblings must survive and reproduce if the genes of an infant are to be fully propagated. In blended families, therefore, offspring should demand approximately four times as much investment as their parents are willing to give, resulting in particularly long and intense periods of parent–offspring conflict. 'Conflict should also be pronounced in families with very young mothers, because younger mothers have more reproductive years ahead of them...' (Simpson 1999, 127). Attachment behaviour would not have evolved if it had only functioned to promote survival through the protection of the child, because survival *per se* is not a goal of natural selection, but differential reproduction. On the one hand, the link between parental sensitivity and the psychological development of children is well established (see Cassidy and Shaver 1999, p. 130–131):

- During the first year of life, insensitive and unresponsive caregiving forecasts the development of insecure attachment (Ainsworth *et al.* 1978).
- Insecurity is associated with a myriad of behaviour problems. As 2 year-olds, insecurely attached children are less tolerant of frustration (Matas, Arend and Scroufe 1978).
- Insecurely attached preschoolers are more likely to be socially withdrawn (Waters *et al.* 1979).
- They are less likely to display sympathy for peers who are upset (Waters, Wippmann and Scroufe 1979).
- They are less willing to interact with friendly adults (Lütkenhaus, Grossmann and Grossmann 1985).
- They are less liked by their classmates (LaFreniere and Sroufe 1985).

On the other hand, however, the question arises as to why, from the perspective of natural history, the whole system should remain so temperamental if it were

so important that the conditions of secure attachment and higher mentalisation be achieved?

The basic functions of the attachment system ought to remain the same over the lifespan (Konner 1982).

> As children move into adolescence, cumulative experiences in relationships are continually assimilated into internal working models... Unlike the attachment system in childhood, however, the system in adulthood becomes integrated with the mating and caregiving systems (Simpson 1999, p. 126; see also Zeifman and Hazan 1997).

One biological explanation as to why distributions of attachment classifications seem similar across cultures is that they represent balanced polymorphisms. A particular morph (e.g. avoidance attachment) is thus only advantageous if it is found in no more than a certain proportion of population. If it becomes too common, its advantages are reduced, and the proportion of individuals of that particular morph declines. If it becomes less common, its characteristics become more advantageous and the proportion possessing it increases. The result is a balanced polymorphism.

DIFFERENCES BETWEEN THE SEXES

It can be assumed that in the course of evolution men and women faced different threats to their survival. For men, such properties as increased aggressiveness, sensation seeking, situational limitations in empathy for war-like arguments, hunting, etc. were perhaps necessary. For women, a mentalisation characterised specifically by empathy perhaps made possible the avoidance of risky situations (rape, etc.) and the choice of an appropriate partner during the EEA. The fact that women make substantially greater initial investments in offspring than men do could be an explanation of the higher mentalisation effort of women and should differentially affect how men and women make reproductive decisions (see Alexander and Noonan 1979; Hinde 1984). At this point the qualification must be added, however, that in the area of attachment styles there are, interestingly enough, hardly any significant sex-related differences.

OTHER MODELS OF BORDERLINE DISORDERS

Next to the mentalisation model described here, there is, of course, a list of important, alternative explanatory models that appear partially consistent with the mentalisation model we are principally investigating here:

- *Model I*. BPD is understood (psychodynamically) as a so-called structural disorder in the area of the stability and maturity of the organisation of personality. This means that increased signs of an 'ego-weakness' and often a 'superego-pathology', as well as a mainly primitive defence mechanism, turn up more often among BPD patients than among normal or 'neurotic' character structures. Central and clinically relevant are identity diffusion and the splitting phenomenon. The splitting serves to prevent (still more) confusion by way of the pseudo-clarity of a 'good and evil dichotomy'. Distinctive behavioural features and relationship problems of the patients are ultimately epiphenomena of the basal (destructive) intra-psychic processes. The increasing disintegration through this (also conflict-full) splitting phenomenon (and its destructive split-off aspects of self and object representations) takes place through the interpretation of this tendency through its re-enactment and through transference (the leading exponent of this model is Otto F. Kernberg). This model would connect to the mentalisation model through the fact that the identity diffusion and the splitting process trace themselves back to mentalisation problems or, vice versa, lead to mentalisation problems. A role in subsequent personality organisation is also played by the fusion-like experienced 'high-peak-affect states'. It is suspected that borderline patients experience an excess of such high-peak-affect states (fear of abandonment, beating, sexual stimulation).
- *Model II*. BPD is understood as a (mainly biologically conditioned) disorder of emotion regulation. Phenomena such as increased impulsiveness, restlessness, risk-taking behaviour or a significant intolerance to frustration are hereby taken into account. These difficulties and the deficit of (life-history validating) alternative skills and corrective experiences lead to the result that the patients are only able to free themselves from these stress situations through often chronic, extremely dysfunctional behaviour patterns (such as self-injury), thus representing attempted solutions, which in turn often act as negative reinforcements. The patient, therefore, has to become an expert of his/her own disorder, whose processes and automatisms (e.g. through chains of behaviour) are 'explained' to him/her during the therapy. This means that the comprehensible aspects ('healthy') of his/her reactions are placed in the centre of the therapy, just like the over-generalised aspects ('pathological') of his/her reactions. He/she thus trains him/herself for alternatives (often experienced almost as aversive by the patient). Actions and feeling are thereby brought into a dialectic relationship. (The leading proponent of this model is Marsha M. Linehan.) Here, the connection to the mentalisation model would be found in the idea that a secure attachment, as well as a higher reflective function ability, could be related to affect differentiation and affect tolerance. It is also possible that the permanent condition of affect dysregulation could lead to mentalisation problems.

- *Model III*. BPD is seen as a (chronic, complex) post-traumatic stress syndrome. To this idea belongs the diagnosis of typical post-traumatic symptoms: dissociative phenomena (e.g. flash-backs, amnesia, hypermnesia, etc.), as well as cognitive retardations, as in the worst forms of neglect. Identity disorders are understood, in this approach, likewise as dissociative mechanisms; in other words, as the sudden appearance (e.g. triggered through stress) of non-integrated partial personalities. The treatment of the disorder proceeds, therefore, through the exposition (to the trauma) on the one hand, but more by working with the split-off aspects of self and object representations, which here are also called 'alters'. The highly dissociative patient, such as one with a dissociated identity disorder, 'gets to know better' all the split-off parts ('personalities') exploratively through imagination-centred therapy, including some which are most frightening. He/she thereby comes to accept their demands and functions. (The leading proponents of this model are Bessel van der Kolk 1994; and Judith Lewis Herman 1992.) Traumatisations can create problems with mentalisation.

Mentalisation theory thus presents the possibility to mediate between these three important models.

The following models of BPD are less important:

- *Model IV*. The borderline disorder is understood as a contemporary form of the earlier hysteria. In this direction could point the following factors: the almost epidemic growth of the disorder in the last few years (as a fashion) and its 'social construction' as an (extreme) form of postmodern identity, with parallels in society (risk-orientated society, the 'flexible person'; Sennett 1998), the 'colourful' picture, the tendency to sexualisation, the sometimes seemingly oedipal dynamic, the partially demonstrative and imitative behaviour (this approach is less widespread). The disorder can hence be sociologically and psychiatric-historically better understood through its demonstrative and constructivistic aspects, as well as with respect to a man–woman dynamic (see e.g. the work of Christopher Bollas 1999 or Ian Hacking 1995).
- *Model V*. There is no qualitative jump between the so-called BPDs and neuroses. It is a matter of severe neurosis, except that the conflicts are more intense and the superego's organisation is more intense. In this sense, BPDs do not exist, only complexly organised, severe neuroses do (this view is especially held by Léon Wurmser 2000).
- *Model VI*. BPD is a spectrum disorder of schizophrenia (a base disorder). According to Kurt Schneider (1959), hallucination and first-level symptoms point in this direction. Through the deepened understanding of severe dissociative disorders (including dissociative identity disorders) and the dividing up by Spitzer, Endicott and Gibbon (1979) of the earlier notion of borderline schizophrenia, latent schizophrenia, pseudoneurotic

schizophrenia, into two nosologically different disorder types: the schizotype personality disorder or schizotypal disorder (ICD-10), this model appears largely obsolete. In the beginning phases of psychoses, however, pseudo-borderline-like clinical presentations do appear (this model is especially represented in Germany by G. Gross and G. Huber 1985).

AMBIVALENT OR RESISTANT ATTACHMENT STYLE AND BORDERLINE PERSONALITY DISORDER

Ambivalent or resistant insecure behaviour is one of the typical groups of the 'strange situation' classification. Children are visibly distressed upon entering the room and often fretful or passive. During separation, they are unsettled or distressed. After reunion, they may alternate bids for contact with signs of angry rejection, such as tantrums, or may appear passive or too upset to signal or make contact. They fail to find comfort in the parent.

Ambivalent children have caregivers who behave in an inconsistent manner (Ainsworth *et al*. 1978), perhaps due to deficient parenting skills. Caregivers of ambivalent infants tend to respond erratically to the needs and signals of their infants, often appearing to be under-involved as parents (Belsky, Rovine and Taylor 1984; Smith and Pederson 1988). Among children who are maltreated, ambivalent children are more likely to have been the victims of parental neglect (Youngblade and Belsky 1989).

> The vehement protest and demanding nature of ambivalent children may therefore reflect an ecologically contingent strategy designed to obtain, retain, and improve the amount of attention and quality of responsiveness from habitually inattentive caregivers (Cassidy and Berlin 1994; Main and Solomon 1986). In other words, the constellation of behaviors characteristic of ambivalent children may have evolved to redress deficiencies in caregiving by young, naïve, overburdened and/or underinvolved parents. For children with such [ambivalent] parents, this behavioral strategy should have permitted greater proximity to the caregivers, solicited better care and increased the children's chances of survival (Simpson 1999, p. 125).

Being an ambivalent and resistant child — with a lack of self-calming strategies and self-reflective functioning — could be interpreted as a consequent survival strategy, a solution to the problem of having an ambivalent, resistant mother, herself lacking self-calming strategies and self-reflective functioning.

In the adult attachment interview, this behaviour seems to be related to the preoccupied (E) attachment representation. Interviewees are placed in the E category when the transcripts suggest an excessive, confused and non-objective preoccupation with particular attachment relationships or experiences.

Discussions of these experiences often appear neither fruitful, objective, nor incisive. Descriptions of early relationships may seem vague and uncritical, or else angry, conflicted and unconvincingly analytical. In studies with BPD patients, most interviewees are placed in the preoccupied category (subtype E3; Fonagy *et al*. 1995).

BELSKY'S MODEL: PATTERNS OF ATTACHMENT AS REPRODUCTIVE STRATEGIES

Belsky *et al*. (1991) have developed a comprehensive, evolution-based lifespan model of human social development (Figure 17.1 outlines the basic features of this model). They contend that the principal evolutionary function of early social experience, moderated by the theory of mind of the caregivers, is to provide children with diagnostic information about the kinds of social and physical environments they are most likely to encounter during their lifetimes. This information should permit individuals to facultatively adopt an appropriate reproductive (and mentalising?) strategy — one that should increase inclusive fitness — in future environments.

As Simpson (1999, p. 130) explains, the model describes two developmental trajectories, culminating in two reproductive strategies in adulthood:

1. One strategy involves a short-term, opportunistic orientation toward mating and parenting, in which sexual intercourse with multiple partners occurs earlier in life, pair bonds are brief and relatively unstable, and parental investment is lower. This orientation is geared toward increasing the quantity of offspring.
2. The alternative strategy involves a long-term investment orientation, in which sexual intercourse occurs later in life with fewer partners, pair bonds are long-term and more stable, and parental investment is greater. This orientation focuses on maximising the quality of offspring.

It is the case, however, that many borderline patients would fit in the first model, but nevertheless exhibit increased sexual difficulties (also inhibitions).

Simpson (1999, p. 132) describes a prototypical situation for persons with BPDs:

> In sum, the Belsky *et al*. (1991) model proposes that early environmental factors that heighten stress should promote harsh, rejecting, or insensitive styles of parenting, which generate insecure working models and patterns of attachment in young children. These models in turn should accentuate and confirm the tenuous, unpredictable nature of close relationships, leading to an opportunistic interpersonal orientation characterized by internalizing or externalizing behavioral disorders. These cues should accelerate sexual maturation, culminating in a short-term reproductive strategy geared toward early reproduction and less parental investment.

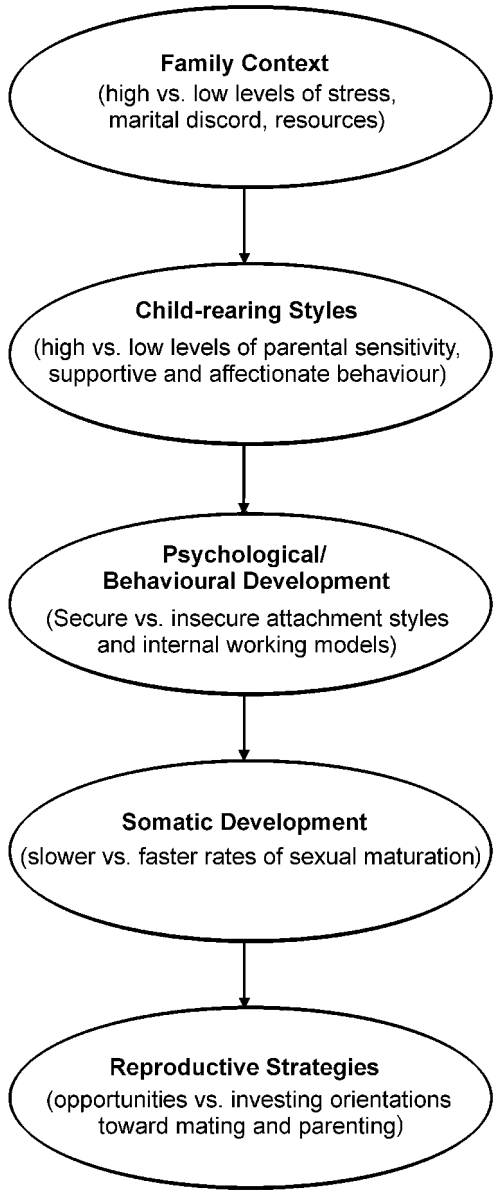


Figure 17.1. Developmental stages leading to opportunistic vs. investing reproductive strategies (based on the model of Belsky *et al.* 1991; Simpson 1999, p. 132)

The model touches on a central hypothesis which, although still unclear, suggests that earlier experienced stress, such as one finds in borderline patients, leads to an earlier onset of puberty (menarche). Furthermore, neglected children indeed appear to become sexually active at an earlier age (Jessor *et al.* 1983; Newcomer and Udry 1987; Moffit *et al.* 1992; Graber, Brooks-Gunn and Warren 1995).

Chisholm's (1996) life-history predictions for ambivalent individuals constitute a distinct departure from Belsky's model. Chisholm contends that ambivalent children should channel greater effort toward early sexual maturity while striving to extract greater investment from their impoverished or under-involved caregivers.

> From the perspective of ultimate causation, this might explain why ambivalent children are so irritable, demanding, and/or preoccupied about gaining and maintaining attention from their caregivers (see Belsky and Cassidy 1994; Kunce and Shaver 1994) (Simpson 1999, p. 134).

As described, this behaviour recalls strongly the demanding and simultaneously irritable manner found in contact with BPD patients. According to Chisholm's model, ambivalent adults should engage in short-term mating, be willing to invest in their children but less able to do so, devote considerable time and energy to parenting, and behave inconsistently toward their children. There are some findings indicating that preoccupied mothers appear especially orientated and responsive toward expressions of fear in their babies (Haft and Slade 1989).

CRITICISM OF THE ONE-SIDED ADAPTIONIST MODEL OF HIGHER MENTALISATION

As Belsky (1999, p. 141) explains, it is not uncommon to read in child and human development textbooks and in scholarly publications that attachment behaviour evolved because it protected infants from predators, and in so doing promoted the survival of the species. Such arguments are insufficient to account for the evolution of attachment and mentalisation behaviour. First, evolution works at the levels of the gene and the individual, not the species. Second, evolution is about differential reproduction, not just survival. Several theorists in this area have understood attachment security and metacognitive abilities as 'overlapping constructs' (Fonagy, Target and Gergely 2000), since the reliability and security of an object relationship permits the child to experience the manifestation of feelings from the other and thus makes the formation of a theory of mind at all possible in the first place.

In contrast to a one-sided overvaluation of an evolutionarily desirable, high degree of mentalisation (as the normal or ideal case of development, as Fonagy

and others assume) as a direct consequence of secure attachment, it could be suggested, in my opinion, that the secure attachment representation does not nearly represent the dominant attachment style (also in Europe, which has been protected for many decades from war and similar catastrophes). Rather, a total of 32 studies have found it only in about 65% of the population in the USA, Europe and Japan (van Ijzendoorn and Sagi 1999, p. 729). It can only be maintained with difficulty that the other 35% of the population are pathological.

> The evolutionary function of the attachment system thus may not be the eliciting of a protective response from a human adult, as Bowlby thought. Rather the survival risks to the organism entailed in the processes of attachment are justified by the benefit that the experience of psychic containment brings in terms of the development of a coherent and symbolising self . . . It is therefore at least plausible to argue that at least one biological function of the process of attachment is the creation of a particular intersubjective environment (Fonagy, 2000b).

Undoubtedly, this perspective, which in the end ascribes adaptive value only to the secure attachment and the high RF, can be brought into question from an evolutionary psychological standpoint. The hypothesis of the social brain connects the development of this ability to certain altered environmental factors, which human beings created or found. (The question in fact arises how the mentalisation ability or the status of reflective functioning presents itself in isolated and still highly 'neolithic' ethnicities, e.g. in Papua New Guinea (Dammann 2002b) or among Australian aborigines.)

With respect to attachment representation I therefore share Belsky's (1999) view that:

> Although it may be the case in contemporary Western society that one pattern, which has come to be known as 'secure', predominates . . . , this should not be taken to mean that in any particular EEA this was the species-typical or normative pattern (Belsky 1999, p. 143).

Since behaviour does not fossilise, there is no way, following Belsky (1999, p. 144), to resolve the theoretically important question of whether we may currently be observing only a subset of attachment and mentalisation patterns that might have once been observed over the course of human evolutionary history, or whether what is observed today reflects the full range of attachment patterns that exist within the human repertoire.

TWO EVOLUTIONARY HYPOTHESES

- *Hypothesis 1*. In contrast to the view proposed especially by Fonagy and his associates, that mentalisation disorders (especially in reflective functioning), as one finds them particularly in borderline disorders, lead almost only to

disastrous consequences, it can be argued with good reason that, also because the system would, in evolutionary psychological terms, not otherwise be so susceptible to disturbances, the possibility of relevant restrictions on mentalisation is ontogenetically as well as phylogenetically adaptive and therefore lies in balance.

Hinde (1986) has suggested that if maternal rejection is induced by harsh environments in which competition for limited resources is intense, offspring who are aggressive and uncooperative may have higher reproductive fitness as adults than offspring without these attributes. Conversely, offspring raised in less hostile environments with more abundant resources should increase their fitness by developing a cooperative and communal orientation toward others in adulthood (Simpson 1999, p. 130). Of course, humans have evolved psychological mechanisms 'capable of ascertaining the degree to which infants were valuable from a reproductive standpoint' (Simpson 1999, p. 129). This is also the reason why, for example, stepfathers are able to treat even sometimes very difficult stepchildren lovingly and sensitively. Possibly, it is this side of empathic mentalisation that enables this response. On the other hand, not all offspring were equal in terms of their reproductive value to parents during evolutionary history. Perhaps, in the balance, the side of the 'refusal of mentalisation' stands for this possibility.

- *Hypothesis 2*. Mentalisation disorders such as one finds in BPDs and disorders marked by a heavily dysfunctional hypersensitivity to empathy represent the extreme opposite to mentalisation disorders such as one finds in autism, which are characterised by a dysfunctional hyposensitivity to empathy. Analogous to the mind-blindness of autism understood by Baron-Cohen as an extreme variant of the male disposition, they are the extreme variant of the female brain or mind.

In their evolutionary theory, Baron-Cohen and Hammer (1997) connect the male brain to the typical extreme form of autistic mentalisation, i.e. large weaknesses in empathy or social intelligence and great strengths in analytical and technical systemisation, especially with respect to inanimate objects (cars, etc.). In another study by this research group (Connellan *et al*. 2000), 102 human neonates (who by definition have not yet been influenced by social and cultural factors) were tested to see whether there was a difference in gazing time at faces (social object) and at mobiles (physical–mechanical object). The results showed that the male infants showed a stronger interest in the physical–mechanical mobiles, while female infants showed a stronger interest in faces.

Analogous to these results, one could describe mentalisation disorders such as those found in BPD patients (70–90% of which are women) as an extreme form of the female brain. These patients show the strongest levels of emotionality and sensibility (proneness to illness, irritability, the tendency to

take everything personally, difficulties in separating themselves from the problems of others), without showing the ability to dissociate themselves (systematically).

INSECURE ATTACHMENT AND LACK OF REFLECTIVE FUNCTIONING: AN EVOLUTIONARY VIEW

> Although caregivers seem prepared to bond easily and strongly with their biological offspring almost from birth, certain conditions can mitigate bonding, resulting in discriminative parental solicitude (Simpson 1999, p. 122).

According to Daly (1989), bonding tends to be hampered (or parental investment is lower) when:

1. One or both parents are not biological relatives.
2. The father's paternity is uncertain.
3. A child is deformed or appears weak.
4. Poverty, lack of food (periods of famine) or too many children in the family reduce the chances of long-term survival.
5. Mothers are very young (even when financial resources and marital status are held constant).
6. Birth spacing is too short.

Spouses who display the lowest levels of support more often have infants with the most severe form of insecurity — the disorganised/disorientated attachment pattern (Spieker 1988; Spieker and Booth 1988). But other studies have found no clear correlation between social support and attachment security (Crockenberg 1981).

> It remains unclear whether patterns of 'disorganised' attachment can be conceptualised in terms of reproductive strategy (Belsky 1999, p. 142).

Step-parents are many times more likely to kill their biologically unrelated stepchildren than are biological parents (Daly and Wilson 1988). Of course, stepfather–stepdaughter 'incest' (if the girl is not a child but has reached reproductive phase) must be reconsidered as an attempt to reproduce genes in a highly complicated 'parent–offspring' situation.

According to a radical sociobiological viewpoint (Daly and Wilson 1981), neonatal emotional bonding for children with congenital handicaps or for mothers who are overburdened (lack of support, depression, etc.) may be disrupted by limited mother–infant post-partum contact or by the inability of mothers and infants to communicate in ways known to facilitate early bonding (e.g. reflective functioning). This finding could also contribute to an explanation as to why the same person can be loving and sensitive to one

child and withdrawn and 'cold' to another (who may have been born under totally different circumstances).

> If precipitating conditions persist over time, parents should continue to engage in insensitive caregiving, with more precipitating conditions resulting in greater insensitivity (Simpson 1999, p. 128).

Post-partum dysphoria and depression could be thus interpreted socio-biologically as an archaic vestige, in the sense of a willingness toward infanticide or, by contrast, toward the responsibility of bonding (Dammann *et al*. 2001, Dammann and Schiefenhövel in preparation).

Various social animal species have been noted to inhibit aggressive attacks when a conspecific displays submission cues. Blair *et al*. (1996) have suggested that humans possess a functionally similar mechanism which mediates the suppression of aggression in the context of distress cues. They have proposed that this mechanism is a prerequisite for the development of the moral/conventional distinction; the consistently observed distinction in the subject's judgement between moral and conventional transgressions. Persons with antisocial or borderline personality disorders may lack this violence inhibitor.

More recently (Mealey 1995; see also Mealey and Kinner in this volume), evolutionary and game theoretic models have tried to present an ultimate explanation of social pathology as the expression of a frequency-dependent life strategy which is selected in dynamic equilibrium in response to certain varying environmental circumstances. But, of course, antisocial strategies (e.g. socio-biological Machiavellianism) are not restricted to sociopaths.

Mealey (1992) has

> argued that the common assumption that an empathy-based approach to predicting the behaviour of others is better than a statistical approach is not necessarily correct; this belief may itself be an emotion-based cognitive bias. To have such a bias may be beneficial, however, for the same reason that emotional commitment biases are beneficial: in situations where voluntary, long-term coalitions can be formed, the personal, empathising (and idealistic) low Machs might outperform the more impersonal, cognitive (and realistic) high Machs, since low Machs would be more successful than high Machs in selecting a cooperator as a partner (Mealey 1995).

I share Fonagy's view that

> the abandonment of reflective function may be seen as constitutional or as an extreme defensive response of children confronted with traumatic situations where they might find overwhelming the contemplation of mental states in their caregiver or themselves (Fonagy 2001, p. 189).

However, when Fonagy continues, 'they thus voluntarily abandon this crucial psychological capacity, with sometimes disastrous consequences', then not only is the 'voluntarily' crucial to recognise, but also the functionality of this inhibition. There is no compelling evolutionary psychological reason for any

pattern (e.g. the secure attachment pattern) to be more primary, more natural, 'healthier' or more specific-typical than any other.

Belsky (1999, p. 144) criticises the stance of several attachment theorists (and Fonagy must also be named among them), which returns to the research by Sroufe (1979). Central to this view is the evolutionarily untenable notion that human beings' natural state is one of grace — to trust others and care for them, giving as much as receiving, if not more so. Only when barriers to this natural state exist will the caring, stress-resilient course of human development be sidetracked, leading to mistrust/insecurity, to dependency, to problematic relations with others or poor mental health.

In contrast, the cross-cultural existence of different attachment patterns and their disputed high stability (see Grossmann, Grossmann and Zimmermann 1999) points rather in the direction that all attachment representation patterns have functional or adaptive aspects. For mentalisation, however, there are hardly any transcultural studies.

> Because there is virtually no acknowledgment that it is not always in a parent's best interest to provide sensitive care (Daly and Wilson 1980; Hrdy 1995). . . . More unappreciated by developmentalists, is the idea that what is in the biological best interest of the parent is not always in the biological best interest of the child (and vice versa)' (Belsky 1999, p. 144).

Sensitivity, secure attachment styles or high degrees of mentalised reflective functioning did not themselves have reproductive-fitness payoffs, at least in certain environments. However, it is the fact that a lesser degree of adaptive functionality can be suspected for the disorganised/detached pattern than for other, likewise insecure attachment styles. The self psychologist Beatrice Beebe has been able to prove empirically that disorganised attachment patterns also appear among extremely considerate mothers, who were hardly able to leave the child alone — perhaps, one could say, because of their own neediness. Tronick, Cohn and Shea (1986) were able to prove in just one experiment that when mothers were asked to put on an expressionless face (still-face) the child's behavioural organisation collapsed.

SEX AND MENTALISATION THEORY

Women appear to have at their disposal a higher ability for mentalisation (empathy toward the other) than men, whose mental activity is characterised by systematic, abstract thinking (of course, it is a matter here merely of a somewhat stronger emphasis). According to Baron-Cohen, the thinking and mentalisation difficulties of persons with autism (who are mostly male) correspond to an almost extreme variation of the 'male brain'. Baron-Cohen (1995) speaks of 'mind-blindness' in connection with autism. This perspective

could not be readily confirmed in Fonagy's studies. With respect to reflective functioning, this means that there is no systematic average advantage for women over men, although it appears to be the case that with men there was a more extreme distribution.

TRANSGENERATIONAL TRANSMISSION

In a large London study (Fonagy, Steele and Steele 1991) carried out on the attachment patterns of future fathers and pregnant women in their last trimester, the attachment style of the mother (whether secure or insecure) proved in 80% of the cases to be predictive of the secure or insecure attachment style of the child (1.5 years later). Sensitive attention to the child explained only 7% of the future secure attachment. Fonagy suspects that metacognitive monitoring (Main 1991) plays an essential role in the explanation of these transgenerational processes, which could close the transmission gap. In sum, the relevance of very high or very low reflective functioning values in the mother plays a large role for the development of the secure attachment style of the child only in groups with significant experiences of deprivation. Put another way, the especially adaptive value of reflective functioning appears to play a specific role in, or become visible through, secure attachment and the maintenance of psychic and traumatic experience only in cases of considerable deprivation.

Countless findings, however, remain in need of explanation. In studies by Fonagy (1996), the early adaptation to the internal working model of the father was significantly more important to an unproblematic development of the child than the adaptation to the internal working model of the mother. Paradoxically, children whose parents are both insecure have fewer adaptation disorders than children with just one insecure parent. For the interpretation of most studies, it also remains problematic that the attachment style and the mentalisation capacity are usually simultaneously collected, which hinders the interpretation of causal connections. It is also possible that both aspects are caused by a superordinate factor and are not causally related at all.

PSYCHOTHERAPEUTIC IMPLICATIONS

In psychiatry and clinical psychology, the mentalisation model has received significant attention for two further reasons, which, however, cannot be further developed here. First, it presents a good model for explaining the transgenerational transmission of pathologies. Second, it stands in immediate proximity to processes of psychotherapy (changes through empathetic, corrective experiences), in which alliance plays an important role. Moreover,

it is the goal of most psychotherapists to further the recognition that outward behaviour is determined by inner conditions.

The task of all psychotherapists is to increase the mentalisation or self-reflective ability of the patient by first of all putting this ability at the patient's disposal, just like a mother would be expected to do. 'The primary task of the therapist consists . . . in transforming inanimate and pre-reflective, teleological concepts of internal states into intentional and animate models' (Daudert 2001, p. 73).

The ability to maintain thinking also in moments of considerable 'brain-storms' (Dennett 1987) presents the therapy of borderline patients with a difficult task:

> The therapist should aim to retain in a part of his or her mind the patient's mental state so as to enable the patient to perceive his or her understanding, notwithstanding the concurrent massive projective processes. The biological pathway, the potential for reflective functioning, remains intact in probably all but the most severely deprived and handicapped [borderline] patients. The experience of intimate contact with another mind capable of recognising the patient's turmoil may be all that is needed for the recovery of a way of being that is essential to adequate functioning in the human world (Fonagy 1998, p. 163).

> The therapist's mentalistic, elaborative stance ultimately enables the [borderline] patient to find himself or herself in the therapist's mind as a thinking, feeling being and to integrate this image as part of his or her sense of himself or herself. . . . The internalisation of the therapist's concern with mental states enhances the patient's capacity for similar concern towards his or her own experience. Respect for minds generates respect for self, respect for others and ultimately respect for the human community (Fonagy, Target and Gergely 2000, p. 117).

From what has been presented here, it is clear that the biggest problem are therapists who, for example, due to their own considerable narcissistic problems, are not themselves prepared to get involved in the other person.

The psychotherapeutic procedure developed by Fonagy (in which the patients are asked to speak in a group about why a fellow patient has just reacted in a certain way) compares to the evolution-inspired procedure suggested by Glantz and Pearce (1989), which emphasises reciprocity in interpersonal relationships (in the sense of reciprocal altruism).

ANXIETY DISORDERS IN BORDERLINE PATIENTS

Anxiety is exceptionally widespread in borderline patients and is considered by some authors to be the principal emotional difficulty they experience. In addition to specific phobias, such as the fear of being abandoned, a chronically diffuse, free-floating anxiety is often described. Aversive behaviours (e.g.

crying, screaming) bring caregivers to children, typically to terminate such actions. Mothers with BPD have difficulties calming their crying children and will even act aggressively in such situations.

As children pass through the toddler years, the desire for physical proximity is slowly replaced by a desire to achieve and maintain psychological or emotional proximity with the parents (Sroufe and Waters 1977). The presence of an attachment figure, later the internalised presence of a good object (Winnicott 1965), should dampen activation of the fear system (sudden movements, darkness, loud noises, being alone). Borderline patients remain often irritable and fearful.

In one study, Adler and Buie (1979) have thus established a connection between the intensive fear of being abandoned found in borderline patients and a structural characteristic of motor-sensory level VI on Piaget's scale of development, to which the notion of object permanence belongs (also, Vygotski's (1978) theory offers points of contact on this issue).

Darwin (1872) had already identified fear as a physiological reaction to potential danger or the occurrence of loss. Fear is universal and found in all higher life-forms. Fear leads to uniform psychological reactions such as 'fight', 'flight', 'freeze' and 'submission'. The human fear reaction also corresponds essentially to these four basic adaptive forms; in human beings, a too-weak as well as too-strong response to fear-causing stimuli. In antisocial personalities and in those who possess an excessive sensation-seeking tendency, the sensation of fear is reduced or even missing. In persons with an anxiety disorder, however, the adaptive fear has made itself independent, appears in an excessive form or in response to inadequate triggers, and becomes thereby maladaptive. Interestingly, in contrast to certain types of antisocial personality disorders and mentalisation difficulties in which a freedom from fear is described, borderline patients demonstrate increased anxiety. The connection between theory of mind and the development of anxiety appears to be not yet totally clarified (on anxiety disorders from an evolutionary psychological perspective, also see e.g. Marks and Nesse 1994).

DEFENCE MECHANISMS IN BORDERLINE PATIENTS IN EVOLUTIONARY PERSPECTIVE

An evolutionary theory of self-deception — the active misrepresentation of reality to the conscious mind — suggests that there may be multiple sources of self-deception in our own species, with important interactions between them. Self-deception may serve to improve deceptions of others. This may include denial of ongoing deception, self-inflation, ego-biased social theory, false narratives of intention, and a conscious mind that operates via denial and

projection (typical 'primitive' defence mechanisms of borderline patients) to create a self-serving world (see Trivers 2000).

The psychodynamically so-called 'identification with the aggressor' is a typical defence mechanism in hostages (Stockholm syndrome) or abused borderline patients. LeCroy (1998) speculates that abuse suffered as a child leads to over-identification with the abuser; 'Self-deception of this kind would have enabled her [an abused woman] to behave devotedly, as abused children frequently do, and thereby solicit nurture'. The need for closeness among traumatised persons remains constant or increases, because the unsuccessful mentalisation precisely makes it impossible to distance oneself or to be alone. Exactly because the *emotional* closeness has become so unbearable, the *physical* closeness of the abuser will be sought out. This offers an explanation for the re-victimisation tendency and the compulsion for re-enactment among borderline patients.

ALTRUISM AND ANTISOCIAL TRAITS

> 'Altruism (Trivers 1971) and cooperativeness — the 'quid-pro-quo' strategy of helping non-kin if, and only if, they have done something for one — might also be underpinned by the mechanism of attachment. Attachment is likely to minimise the adverse effects of 'cheaters' — individuals who do not reciprocate equitably in groups over time and to whom we are unlikely to become attached' (Fonagy *et al.* 2002, p. 122).

Trivers unveiled the theory of parental investment and sexual selection and later he introduced the theory of parent–offspring conflict. This theory recognised [in a direct extension of Hamilton's (1964) theory of kin selection] that children (who share only 50% of their genes with parents and full siblings) should desire greater investment than their parents have been selected to provide.

Connected to this idea is the fact that borderline patients who are themselves given up to a care institution or a foster home, in other words whose parents rescinded their investment in the child early on, often treat their own children in the same manner, although they themselves experienced the negative aspects of this treatment and could therefore be expected to spare their own children such suffering.

> Although they may not yield developmental outcomes that a mental health perspective values or that our society wants to promote, contemporary (or even enduring) cultural values should not be confused with biological desiderata (Hinde and Stevenson-Hinde 1990).

From a sociobiological perspective, this argument is, in my view, to be taken seriously. From a mental health perspective it could, however, be objected that, as a rule, patients with severe personality disorders exhibit significant

difficulties in countless areas of life (relationship problems, increased rate of suicide, etc.) which typically considerably impair their reproductive fitness and also, in most conceivable environments, would scarcely appear to be adaptive. The quantity-vs.-quality orientation toward reproduction seems not always to be evident among borderline patients, who often do not have children because of their own significant difficulties. But, 'Many animals, ranging from eusocial insects (e.g. wasps, bees, ants) to birds (e.g. acorn woodpeckers) and a few mammals (e.g. naked mole rats) produce offspring that are physically or behaviourally sterile (Baker and Bellis 1995)' (Belsky 1999, p. 156). Such seemingly non-reproductive behaviour has even been observed among humans (childless women as sterile helpers to their parents and siblings are indirectly reproductive 'helper-at-the-nest-behaviour') (Clark 1996; Borgerhoof Mulder 1992).

In a recent investigation, van Ijzendoorn *et al.* (1997) examined the relationship between attachment representations and personality disorders in a sample of 40 young men held for the commission of serious crimes. Using the AAI and the Structured Interview for Disorders of Personality (Pfohl *et al.* 1989), these researchers demonstrated that secure attachment representations were virtually absent in the sample; separation from attachment figures in childhood was related to current insecure attachment, as well as to personality disorders. In particular, insecure attachment was associated with disturbances such as narcissistic, sadistic and antisocial personality disorders (van Ijzendoorn *et al.* 1997).

Similarly, Fonagy *et al.* (1996) and Fonagy and Target (1995) compared the attachment representations of offenders who had committed crimes against property and those who had committed violent crimes, such as rape and murder. The latter group reported more extremely disturbed attachment representations, often accompanied by a history of abuse. In parallel, they also failed to demonstrate ability to reflect on and take into consideration the mental lives and emotions of others.

> Taken together, these investigations serve to highlight the advances made by researchers in empirically studying the attachment representations of violent criminal offenders who, beyond their severe antisocial behaviours, appear unaffected by the pain of others (Saltaris 2002, p. 741). Several studies document a relation between insecure or disorganised attachment patterns and childhood aggression or conduct disorders (Lyons-Ruth, Alpern and Repecholi 1993).

DISCUSSION

From an evolutionary psychological perspective, one could in summary establish, according to the hypotheses proposed here, that with regard to the development and transmission of mentalisation abilities (and, in this respect,

especially reflective functioning) two strategies could be chosen, which could contain adaptive as well as maladaptive aspects.

The first strategy means the transmission and further development of the ability to mentalise (or, otherwise stated, one could here also speak of a progressive theory of mind). This strategy can be seen as largely adaptive, especially in social environments characterised by the following ecological–psychological factors:

- Low population density.
- Good availability of nutritional resources (low competition).
- Low levels of violent conflict.

The second strategy means the inhibition or the retraction of the ability to mentalise (one could speak here of a defensive or withdrawn theory of mind). This strategy then can contain a renewed adaptive functioning (with reproductive advantages), when the environment is characterised by the following factors:

- High population density (crowding).
- Deficit of nutritional resources (high competition).
- Extensive violent conflict.

Both theoretical models and empirical research (see Mealey 1995) show that in societies that are becoming larger and more competitive, individuals become more anonymous and more Machiavellian, leading to reductions in altruism and increases in crime. Social stratification and segregation or problems in the family can also lead to feelings of inferiority, pessimism and depression among the less-privileged, which can in turn promote the use of alternative competitive strategies, including antisocial or borderline-like behaviour (e.g. without taking on each social responsibility) (Magid and McKelvey 1987; Sanchez 1986).

Generally, this conceptualisation has a certain closeness to the notion of the 'secondary sociopath', as presented by Mealey (1995), who wrote:

> Secondary sociopaths are individuals who use an environmentally contingent, facultative cheating strategy not as clearly tied to genotype; this strategy develops in response to social and environmental conditions related to disadvantage and even within an individual lifetime, with variation in immediate social circumstances.

According to my hypothesis, not only must attachment but also the mentalisation ability stand in a kind of flexible balance, due to changing environmental variables.

> When an evolutionary perspective is applied to temperament, it raises the possibility that infants and children may vary, for heritable reasons, in their susceptibility to environmental influence (Belsky, 1997). Thus it seems plausible to entertain the prospect that for inborn, constitutional/temperamental reasons, some infants may be very strongly predisposed

> to develop secure, avoidant, or resistant attachments almost regardless of the quality of care they experience. . . . (Belsky, 1999, p. 157).

Similarly, Belsky (1999, p. 143) writes:

> Under the diverse conditions in which hominids evolved, it seems more reasonable to presume that no pattern of attachment was primary and others secondary, but rather that what evolved was a repertoire of attachment behaviours that could be flexibly organised into different patterns contingent on ecological and caregiving conditions (Belsky *et al.* 1991; Chisholm 1996; Hinde 1982). That is, in the same way the natural selection shaped the body to store fat under some ecological conditions but not others, so it shaped patterns of attachment (Belsky 1999, p. 143).

I would argue that the same is true for mentalising capacities in humans. Elsewhere, I have indicated that affective disorder could, from an evolutionary-psychological standpoint, be understood in countless ways as a balanced disorder (Dammann 2002a).

Interestingly, the deficit or disorder of reflective functioning does not always appear in borderline patients. Rather, it is specific to relationship contexts in which the limited and insecure internal working models are activated, especially in the case of conflicts with others. 'The borderline patient is not "mind blind", rather, he/she is not "mind conscious" ' (Fonagy 1999). Patients with severe personality disorders can often develop 'a certain level of non-conscious mind-reading skills. Reflective function does not permanently disappear' (Fonagy, Target and Gergely 2000, p. 116). This opens up theoretical questions which are in my opinion still insufficiently explained. How can one represent a partial mentalisation disorder? Does this mean that mentalisation capacities are inseparably bound with object-relation typologies, such as in patients with severe hysterical neuroses who have no problems in the workplace or in their families but suffer from shipwrecks in all love relationships? If so, then the question arises whether in all severe neuroses it is a matter of conflicts or of structural integration deficits.

From an evolutionary-psychological perspective, the idea of a universal structural personality organisation in human beings appears scarcely plausible. From the side of human ethology, the quick change from gentleness and social relatedness to violent aggression and cruelty is impressive in traditional ethnicities (e.g. in Melanesia), whose environment partially corresponds to one of several hypothetical 'environments of evolutionary adaptedness'.

Families that are shaped by break-ups, violence, separation or sexual abuse could exhibit, so to speak, a micro-milieu of an anomalous civil war situation. In the meantime, the damaging influence on reflective functioning of traumatisations such as physical or sexual abuse has been well proved (Beeghly and Cicchetti 1994; Fonagy 2000).

Cicchetti (1989) has been able to prove that traumatised children avoid empathising in the psychological condition of others, e.g. out of fear that they

will discover hostility, aggressiveness or rejection in the thoughts and desires of their caregivers, in order to protect themselves from further pain. 'The child may turn away from the world of minds altogether and refuse to conceive of the mental states of his or her attachment objects' (Fonagy 1998, p. 161). In my opinion, this individual avoidance strategy corresponds to a species-specific behavioural option. It therefore appears understandable that mentalisation ability, which probably developed rather late in human history, must still be subject to a certain dynamic flexibility.

ACKNOWLEDGEMENTS

I would like to thank Susanne Brunhart Wiggins BA, Germantown, MD, who read and translated parts of an earlier version of this chapter.

REFERENCES

Adler, G. and Buie, D.H. (1979). Aloneness and borderline psychopathology: the possible relevance of child development issues. *Int. J. Psychoanal.* **60**, 83–96.

Ainsworth, M.D., Blehar, M.C., Waters, E., and Wall, S. (1978). *Patterns of Attachment: a Psychological Study of the Strange Situation*. Erlbaum, Hillsdale, NJ.

Alexander, R.D. and Noonan, K.M. (1979). Concealment of ovulation, parental care, and human social evolution. In N.A. Chagnon and W.W. Irons (eds), *Evolutionary Biology and Human Social Behavior*. Duxbury, North Scituate, MA, pp. 402–435.

Baker, R. and Bellis, M. (1995). *Human Sperm Competition*. Chapman and Hall, New York.

Barkow, J.H., Cosmides, L. and Tooby, J. (eds) (1992). *The Adapted Mind. Evolutionary Psychology and the Generation of Culture*. Oxford University Press, New York.

Baron-Cohen, S. (1995). *Mindblindness. An Essay on Autism and Theory of Mind*. MIT Press, Cambridge, MA.

Baron-Cohen, S. and Hammer J. (1997). Is autism an extreme form of the male brain? *Adv. Infancy Res.* **11**, 193–217.

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

Beebe, B. and Lachmann, F. (1988). The contribution of mother–infant mutual influence to the origins of self and objects representations. *Psychoanal. Psychol.* **5**, 305–337.

Beeghly, M. and Cicchetti, D. (1994). Child maltreatment, attachment, and the self system: emergence of an internal state lexicon in toddlers at high social risk. *Dev. Psychopathol.* **6**, 5–30.

Belsky, J. and Cassidy, J. (1994). Attachment: theory and evidence. In M.L. Rutter, D.F. Hay and S. Baron-Cohen (eds), *Development Through Life: A Handbook for Clinicians*. Blackwell, Oxford, pp. 373–402.

Belsky, J., Rovine, M. and Taylor, D.G. (1984). The Pennsylvania Infant and Family Development Project. III: The origins of individual differences in infant–mother attachment: maternal and infant contributions. *Child Dev.* **55**, 718–728.

Belsky, J. (1997). Variation in susceptibility to environmental influence: an evolutionary argument. *Psychol. Inq.* **8**, 82–186.

Belsky, J. (1999). Modern evolutionary theory and patterns of attachment. In J. Cassidy and P.R. Shaver (eds), *Handbook of Attachment: Theory, Research, and Clinical Applications*. Guilford, New York, pp. 141–161.

Belsky, J., Steinberg, L. and Draper, P. (1991). Childhood experience, interpersonal development, and reproductive strategy: an evolutionary theory of socialization. *Child Dev.* **55**, 718–728.

Belsky, J., Rosenberger, K. and Crnic, K. (1995). The origins of attachment security: 'classical' and contextual determinants. In S. Goldberg, R. Muir and J. Kerr (eds), *Attachment Theory: Social, Developmental and Clinical Perspectives*. Analytic Press, Hillsdale, NJ, pp. 153–183.

Bion, W.R. (1962). *Learning from Experience*. Heinemann, London.

Blair, R.J.R., Sellars, C., Strikland, I., Clark, F. *et al.* (1996). Theory of mind in the psychopath. *J. Forens. Psychiat.* **7**, 15–25.

Bolas, C. (1999). *Hysteria*. Routledge, London and New York.

Borgerhoff Mulder, M. (1992). Reproductive decisions. In E.A. Smith and B. Winterholder (eds), *Evolutionary Ecology and Human Behavior*. Aldine de Gruyter, New York, pp. 129–147.

Bowlby, J. (1973). *Attachment and Loss: Vol. 2. Separation*. Basic Books, New York.

Bowlby, J. (1980). *Attachment and Loss: Vol. 3. Loss*. Basic Books, New York.

Bretherton, I. (1991). Pouring new wine into old bottles: the social self as internal working model. In M.R. Gunnar and L.A. Sroufe (eds), *Self Processes and Development*. Minnesota Symposia on Child Psychology, vol. 23. Erlbaum, Hillsdale, NJ, pp. 1–41.

Bretherton, I. and Mulholland, K.A. (1999). Internal working models in attachment relationships: a construct revisited. In J. Cassidy and P.R. Shaver (eds), *Handbook of Attachment: Theory, Research and Clinical Applications*. Guilford, New York, pp. 89–114.

Brewer, M.B. and Caporael, L.R. (1990). Selfish genes vs. selfish people: sociobiology as origin myth. *Motivation Emotion* **14**, 237–243.

Brothers, L. (1990). The social brain: a project for integrating primate behavior and neurophysiology in a new domain. *Concepts Neurosci.* **1**, 27–51.

Brüne, M. (2001). Social cognition and psychopathology in an evolutionary perspective — current status and proposals for research. *Psychopathology* **34**, 85–94.

Brüne M. and Ribbert, H. (eds) (2001). *Evolutionsbiologische Konzepte in der Psychiatrie*. Peter Lang, Frankfurt am Main.

Bucci, W. (1997). *Psychoanalysis and Cognitive Science: A Multiple Code Theory*. Guilford, New York.

Cassidy, J. and Berlin, L. (1994). The insecure/ambivalent pattern of attachment: theory and research. *Child Development* **65**, 971–991.

Cassidy, J. and Shaver, P.R. (eds) (1999). *Handbook of Attachment: Theory, Research, and Clinical Applications*. Guilford, New York.

Chisholm, J. (1996). The evolutionary ecology of attachment organization. *Hum. Nature* **7**, 1–38.

Christie, R. and Geis, F.L. (1970). *Studies in Machiavellianism*. Academic Press, New York.

Cicchetti, D. and Lynch, M. (1995). Failures in the expectable environment and their impact on individual development: the case of child maltreatment. In D. Cicchetti and D.J. Cohen (eds), *Developmental Psychopathology, vol. 2: Risk, Disorder, and Adaptation*. Wiley, New York, pp. 32–71.

Cicchetti, D. (1989). How research on child maltreatment has informed the study of child development: perspectives from developmental psychopathology. In D. Cicchetti and V. Carlson (eds), *Child Maltreatment: Theory and Research on the Causes and Consequences of Child Abuse and Neglect*. Cambridge University Press, Cambridge, pp. 377–431.

Clark, C. (1996). Reproductive consequences of human dispersal in nineteenth century Sweden. *Behav. Ecol.* **41**, 17–31.

Clarkin, J.F. and Kernberg, O.F. (2002). Psychopathology and change: early findings of our project. Presentation at the 3rd. Annual Meeting Borderline Personality Disorder Research Foundation, Basel, 10–12 July 2002.

Clarkin, J.F., Widiger, T., Frances, A., Hurt, S.W. and Gilmore, M. (1983). Prototypic typology and the borderline personality disorder. *J. Abnorm. Psychol.* **92**, 263–275.

Connellan, J., Baron-Cohen, S., Wheelwright, S. Batki, A. and Ahluwalia, J. (2000). Sex differences in human neonatal social perception. *Infant Behav. Dev.* **23**, 113–118.

Cowdry, R.W., Pickar, D. and Davies, R. (1985). Symptoms and EEG findings in the borderline syndrome. *Int. J. Psychiat. Med.* **15**, 201–211.

Crockenberg, S.B. (1981). Infant irritability, mother responsiveness, and social support influences on the security of infant–mother attachment. *Child Dev.* **52**, 857–869.

Daly, M. and Wilson, M. (1980). Discriminative parental solicitude: a biological perspective. *J. Marriage Fam.* **42**, 277–288.

Daly, M. and Wilson, M.I. (1981). Abuse and neglect of children in evolutionary perspective. In R.D. Alexander and D.W. Tinkle (eds), *Natural Selection and Social Behavior: Recent Research and New Theory*. Blackwell, Oxford, pp. 405–416.

Daly, M. and Wilson, M. (1987). Children as homicide victims. In R.J. Gelles and J.B. Lancaster (eds), *Child Abuse and Neglect: Biosocial Dimensions*. Aldine de Gruyter, New York, pp. 201–204.

Daly, M. and Wilson, M. (1988). *Homicide*. Aldine de Gruyter, New York.

Daly, M. (1989). Parent–offspring conflict and violence in evolutionary perspective. In R.W. Bell and N.J. Bell (eds), *Sociobiology and the Social Sciences*. Texas Tech University Press, Lubbock, TX, pp. 25–43.

Damasio, A.R. (1995). Towards a neurobiology of emotion and feeling: operational concepts and hypotheses. *Neuroscientist* **1**, 19–25.

Dammann, G. (2001). Die Relevanz neuerer Theorien zur Mentalisierung für Verständnis und Therapie der Borderline-Störung. *Nervenheilkunde* **20** (suppl. 3), S28–29.

Dammann, G. (2002a). Depressive Symptome aus der Perspektive der Evolutionären Psychiatrie. *Nervenheilkunde* **21**(3), 132–137.

Dammann, G. (2002b). Das ethnologische 'Big man–Paradigma' — ein Modell für die narzisstische Persönlichkeitsstörung? In Symposium 'Persönlichkeitsstörungen aus evolutionspsychiatrischer Perspektive', G. Dammann, M. Brüne and M.H. Stone (Chairs), 5th International Society for the Study of Personality Disorders European Congress on Personality Disorders, Munich, 5 July 2002.

Dammann, G. and Schiefenhövel, W. (eds) (2003). *Anthropological, Ethnopsychiatrical and Evolutionary Perspectives of Postpartum Affective Disorders*. (in preparation).

Dammann, G., Ackermann, E., Delius, M., Strobl, C. *et al.* (2001). The Munich evolution–psychological postpartum study: hypotheses, design, and first results. *Arch. Women's Ment. Health* **3** (suppl. 2), 82.

Darwin C. (1872). *The Expressions of the Emotions in Man and Animals*. Murray, London.

Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex*. Princeton University Press, Princeton, NJ (reprinted 1981).

Daudert, E. (2001). Selbstreflexivität, Bindung und Psychopathologie. Zusammanhaenge bei stationaeren Gruppenpsychotherapie-Patienten (= Phil. Diss. Hamburg), Kovač, Hamburg.

De Wolff, M.S. and van Ijzendoorn, M.H. (1997). Sensitivity and attachment: a meta-analysis on parental antecedents of infant attachment. *Child Dev*. **68**, 571–591.

Dennett, D.C. (1987). *The Intentional Stance*. MIT Press, Cambridge, MA.

Dunbar, R.I.M. (1998). The social brain hypothesis. *Evol. Anthropol*. **6**, 178–190.

Dunn, J. (1988). *The Beginnings of Social Understanding*. Harvard University Press, Cambridge, MA.

Erickson M.F., Sroufe, L.A. and England, B. (1985). The relationship between quality of attachement and behavior problems in preschool in a high risk sample. In I. Bretherton and E. Waters (eds), *Growing Points of Attachment Theory and Research. Monogr. Soc. Res. Child Dev*. **50** (1–2, Serial No. 209), 147–193.

Fonagy, P. (1989). On tolerating mental states: theory of mind in borderline patients. *Bull. Anna Freud Centre* **12**, 91–115.

Fonagy, P. (1991). Thinking about thinking: some clinical and theoretical considerations in the treatment of a borderline patient. *Int. J. Psychoanal*. **72**, 1–18.

Fonagy, P. (1996). The significance of the development of metacognitive control over mental representations in parenting and infant development. *J. Clin. Psychoanal*. **5**, 67–86.

Fonagy, P. (1997). Attachment and reflective function: their role in self-organization. *Dev. Psychopathol*. **9**, 679–700.

Fonagy, P. (1998). Attachment theory approach to treatment of the difficult patient. *Bull. Menninger Clin*. **62**, 147–169.

Fonagy, P. (1999). Pathological attachments and therapeutic action. Paper to the Developmental and Psychoanalytic Discussion Group, American Psychoanalytic Association Meeting, Washington, DC, 13 May 1999.

Fonagy, P. (2000). Attachment in infancy and the problem of conduct disorders in adolescence: the role of reflective function. Plenary address to the International Association of Adolescent Psychiatry, San Francisco, January 2000.

Fonagy, P. (2000b). Points of contact and divergence between psychoanalytic and attachment theories: is psychoanalytic theory truly different? Paper presented as visiting Professor of Psychoanalysis of the Michigan Psychoanalytic Institute, 2–9 April 2000 (http://www.psychol.ucl.ac.uk/psychoanalysis/zipped/michigan4.rtf).

Fonagy, P. (2001). *Attachment Theory and Psychoanalysis*. Other Press, New York.

Fonagy, P. and Target, M. (1995). Towards understanding violence: the use of the body and the role of the father. *Int. J. Psychoanal*. **76**, 487–502.

Fonagy, P., Steele, H. and Steele, M. (1991). Maternal representations of attachment during pregnancy predict the organization of infant–mother attachment at 1 year of age. *Child Dev*. **62**, 891–905.

Fonagy, P., Steele, M., Steele, H., Leigh, T. *et al*. (1995). Attachment, the reflective self, and borderline states: the predictive specificity of the Adult Attachment Interview and pathological emotional development. In S. Goldberg, R. Muir, J. Kerr (eds), *Attachment Theory: Social, Developmental and Clinical Perspectives*. Analytic Press, New York, pp. 233–278.

Fonagy, P., Leigh, T., Steele, M., Steele, H. *et al*. (1996). The relation of attachment status, psychiatric classification, and response to psychotherapy. *J. Consult. Clin. Psychol*. **64**, 22–31.

Fonagy, P., Target, M. and Gergely, G. (2000). Attachment and borderline personality disorder: a theory and some evidence. *Psychiat. Clin. N. Am*. **23** (1), 103–123.

Fonagy, P., Gergely, G., Jurist, E.L. and Target, M. (2002). *Affect Regulation, Mentalization, and the Development of the Self*. Other Press, New York.
Geis, F. and Levy, M. (1970). The eye of the beholder. In R. Christie and F.L. Geis (eds), *Studies in Machiavellianism*. Academic Press, New York.
George, C., Kaplan, N. and Main, M. (1985). The Attachment Interview for Adults. Unpublished Manuscript, University of California, Berkeley, CA.
Gergely, G. (1994). From self-recognition to theory of mind. In S.T. Parker, R.W. Mitchell and M.L. Boccia (eds), *Self-Awareness in Animals and Humans: Developmental Perspectives*. Cambridge University Press, Cambridge, pp. 51–61.
Glantz, K. and Pearce, J.K. (1989). *Exiles from Eden. Psychotherapy from an Evolutionary Perspective*. Norton, New York.
Gordon, R.M. (1986). Folk psychology as simulation. *Mind Lang*. **1**, 158–171.
Gordon, S.L. (1989). The socialization of children's emotions: emotional culture, competence, and exposure. In C. Saarni and P.L. Harris (eds), *Children's Understanding of Emotion*. Cambridge University Press, Cambridge, pp. 319–349.
Graber, J., Brooks-Gunn, J. and Warren, M. (1995). The antecedents of menarcheal age. *Child Dev*. **66**, 346–359.
Grieser, D.L. and Kuhl, P.K. (1988). Maternal speech to infants in a tonal language: support for universal prosodic features in motherese. *Dev. Psychol*. **24**, 14–20.
Gross, G. and Huber, G. (1985). Psychopathology of basic stages of schizophrenia in view of formal thought disturbances. *Psychopathology* **18**, 115–125.
Grossman, K.E. and Grossman, K. (1991). Attachment quality as an organizer of emotional and behavioural responses in a longitudinal perspective. In C.M. Parkes, J. Stevenson-Hinde and J. Marris (eds), *Attachment Across the Life Cycle*. Routledge, London, pp. 93–114.
Grossmann, K.E., Grossmann, K. and Zimmermann, P. (1999). A wider view of attachment and exploration. In J. Cassidy and P.R. Shaver (eds), *Handbook of Attachment: Theory, Research and Clinical Applications*. Guilford, New York, pp. 760–786.
Hacking, I. (1995). *Rewriting the Soul. Multiple Personality and the Sciences of Memory*. Princeton University Press, Princeton, NJ.
Haft, W. and Slade, A. (1989). Affect attunement and maternal attachment: a pilot study. *Infant Ment. Health J*. **10**, 157–172.
Hamilton, W.D. (1964). The genetical evolution of social behaviour. *J. Theoret. Biol*. **7**, 1–52.
Hare, B., Call, J. and Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behav*. **61**, 139–151.
Hellman, I.D., Morrison, T.L. and Abramowitz, S.I. (1986). The stresses of psychotherapeutic work: a replication and extension. *J. Clin. Psychol*. **42**, 197.
Herman, J.L. (1992). *Trauma and Recovery*. Basic Books, New York.
Hinde, R. (1982). Attachment: Some conceptual and biological issues. In C.M. Parkers and J. Stevenson-Hinde (eds), *The Place of Attachment in Human Behavior*. Basic Books, New York, pp. 14–37.
Hinde, R.A. (1984). Why do the sexes behave differently in close relationships? *J. Soc. Pers. Relationships* **1**, 471–501.
Hinde, R.A. (1986). Some implications of evolutionary theory and comparative data for the study of human prosocial and aggressive behavior. In D. Olweus, J. Block and M. Radke-Yarrow (eds), *Development of Antisocial and Prosocial Behavior*. Academic Press, Orlando, FL, pp. 13–32.
Hinde, R. and Stevenson-Hinde, J. (1990). Attachment: biological, cultural and individual desiderata. *Hum. Dev*. **33**, 62–72.

Hobson, R.P. (1993). *Autism and the Development of Mind*. Erlbaum, London.
Hrdy, S. (1995). Natural-born mothers. *Nat. Hist.* **104**, 32–42.
Jessor, R., Costa, F., Jessor, L. and Donovan, J. (1983). Time of first intercourse: a prospective study. *J. Personality Soc. Psychol.* **44**, 608–626.
Joffe, T.H. (1997). Social pressures have selected for an extended juvenile period in primates. *J. Hum. Evol.* **32**, 593–605.
Kass, F., Skodol, A.E., Charles, E., Spitzer, R.L. and Williams, J.B.W. (1985). Scaled ratings of DSM-III personality disorders. *Am. J. Psychiat.* **142**, 478–483.
Klaus, M. and Kennell, J. (1976). Parent-to-infant attachment. In D. Hull (ed.), *Recent Advances in Pediatrics*. Churchill Livingstone, New York, pp. 129–152.
Konner, M. (1982). *The Tangled Wing: Biological Constraints on the Human Spirit*. Holt, Rinehart, and Winston, New York.
Kosslyn, S.M. (1987). Seeing and imaging in the cerebral hemispheres: a computational approach. *Psychol. Rev.* **94**, 148–175.
Kraemer, G.W. (1992). A psychobiological theory of attachment. *Behav. Brain Sci.* **15** (3), 493–541.
Kunce, L.J. and Shaver P.R. (1994). An attachment-theoretical approach to caregiving in romantic relationships. In K. Bartholomew and D. Perlman (eds), *Attachment Processes in Adulthood*. Jessica Kingsley, London, pp. 205–237.
LaFreniere, P.J. and Sroufe, L.A. (1985). Profiles of peer competence in the preschool: interrelations between measures, influence of social ecology, and relation to attachment history. *Dev. Psychol.* **21**, 56–69.
LeCroy, D. (1998). Darwin in the clinic: an evolutionary perspective on psychodynamics found in a single case. *Across Species Comparison and Psychopathology Newsletter* **11**, 6–12.
Leslie, A. (1987). Pretence and representation: the origins of 'theory of mind'. *Psychol. Rev.* **94**, 412–426.
Lewis, M. and Brooks, J. (1975). Infant's social perception: a constructivist view. In L. Cohen and P. Salapatek (eds), *Infant Perception: From Sensation to Cognition*, vol. 2. Academic Press, New York, pp. 102–148.
Liotti, G. (1992). Disorganized/disoriented attachment in the etiology of the dissociative disorders. *Dissociation* **5**, 196–204.
Lorenz, K. (1973). *Behind the Mirror: a Search for a Natural History of Human Knowledge*. Harcourt Brace Jovanovich, New York.
Lütkenhaus, P., Grossmann, K.E. and Grossmann K. (1985). Infant–mother attachment at twelve months and style of interaction with a stranger at the age of three years. *Child Dev.* **56**, 1538–1542.
Lyons-Ruth, K., Alpern, L. and Repacholi, B. (1993). Disorganized infant attachment classification and maternal psychosocial problems as predictors of hostile–aggressive behavior in the preschool classroom. *Child Dev.* **64**, 572–585.
Magid, K. and McKelvey, C.A. (1987). *High risk: Children without a Conscience*. Bantam Books, New York.
Main, M. (1981). Avoidance in the service of attachment: a working paper. In K. Immelmann, G. Barlow, M. Main and L. Petrinovich (eds), *Behavioral Development: the Bielefeld Interdisciplinary Project*. Cambridge University Press, New York, pp. 651–693.
Main, M. (1990). Cross-cultural studies of attachment organization: Recent studies, changing methodologies, and the concept of conditional strategies. *Hum. Dev.* **33**, 48–61.
Main, M. (1991). Metacognitive knowledge, metacognitive monitoring, and singular (coherent) vs. multiple (incoherent) models of attachment: findings and directions for

future research. In P. Harris, J. Stevenson-Hinde and C. Parkes (eds), *Attachment across the Lifecycle*. Routledge, New York, pp. 127–159.
Main, M. and Goldwyn, R. (1994). Adult Attachment Rating and Classification System. Manual in Draft, Version 6.0. Unpublished manuscript. University of California at Berkeley, Berkeley, CA.
Main, M. and Solomon, J. (1986). Discovery of an insecure-disorganized/disoriented attachment pattern: procedures, findings, and implications for the classification of behavior. In T.B. Brazelton and M. Yogman (eds), *Affective Development in Infancy*. Ablex, Norwood, NJ, pp. 95–124.
Mandler, J. (1992). How to build a baby: II. Conceptual primitives. *Psychol. Rev.* **99**, 587–604.
Marks, I.M. and Nesse, R.M. (1994). Fear and fitness: an evolutionary analysis of anxiety disorders. *Ethol. Sociobiol.* **15**, 247–261.
Matas, L., Arend, R. and Sroufe, L.A. (1978). Continuity in adaptation in the second year: the relationship between quality of attachment and later competence. *Child Dev.* **49**, 547–556.
McGlashan, T.H. (1986). The Chestnut Lodge follow-up study: III. Long-term outcome of borderline personalities. *Arch. Gen. Psychiat.* **43**, 20–30.
Mealey, L. (1992). Are monkeys nomothetic or idiographic? *Behav. Brain Sci.* **15**, 161–162.
Mealey, L. (1995). The sociobiology of sociopathy: an integrated evolutionary model. *Behav. Brain Sci.* **18**, 523–541.
Miller, G. (2000). Mental traits as fitness indicators: expanding evolutionary psychology's adaptionism. In D. LeCroy and P. Moller (eds), *Evolutionary Perspectives on Human Reproductive Behavior. Ann. N. Y. Acad. Sci.* **907**, pp. 62–74.
Moffit, T.E., Caspi, A., Belsky, J. and Silva, P.A. (1992). Childhood experience and the onset of menarche: a test of a sociobiological model. *Child Dev.* **63**, 47–58.
Moore, C. and Frye (1991). The acquisition and utility of theories of mind. In D. Frye and C. Moore (eds), *Children's Theories of Mind. Mental States and Social Understanding*. Erlbaum, Hillsdale, NJ, pp. 1–38.
Moses, L.J. and Chandler M.J. (1992). Traveler's guide to children's theories of mind. *Psychol. Inq.* **3**, 286–301.
Murray, L. and Cooper, P.J. (eds) (1997). *Postpartum Depression and Child Development*. Guilford, New York.
Newcomer, S. and Udry, J. (1987). Parental marital status effects on adolescent sexual behavior. *J. Marriage Fam.* **49**, 235–240.
Pfohl, B., Blum, N., Zimmerman, M. and Stangl, D. (1989). *Structured Interview for DSM-III-R Personality-Revised (SIDP-R)*. Department of Psychiatry, University of Iowa, Ames, IA.
Premack, D. and Woodruff, G. (1978). Does the chimpanzee have a 'theory of mind'? *Behav. Brain Sci.* **4**, 515–526.
Saltaris, C. (2002). Psychopathology in juvenile offenders. *Clin. Psychol. Rev.* **22** (5), 729–752.
Sanchez, J. (1986). Social crises and psychopathy: toward a sociology of the psychopath. In W.H. Reid, D. Dorr, J.I. Walker and J.W. Bronner, III (eds), *Unmasking the Psychopath: Antisocial Personality and Related Syndromes*. W.W. Norton, New York, pp. 78–97.
Schmitt, A. and Grammer, K. (1997). Social intelligence and success: don't be too clever in order to be smart. In A. Whiten, R.W. Byrne (eds), *Machiavellian Intelligence II. Extensions and Evaluations*. Cambridge University Press, Cambridge, pp. 86–111.
Schneider, K. (1959). *Clinical Psychopathology*. Grune and Stratton, London.

Schore, A. (1994). *Affect Regulation and the Origin of the Self: the Neurobiology of Emotional Development*. Erlbaum, Mahwah, NJ.
Sennett, R. (1998). *The Corrosion of Character: the Personal Consequences of Work in the New Capitalism*. W.W. Norton, New York.
Shaver, P., Collins, N. and Clark, C. (1996). Attachment styles and internal working models of self and relationship partners. In G. Fletcher and J. Fitness (eds), *Knowledge Structures in Close Relationships*. Erlbaum, Mahwah, NJ, pp. 25–62.
Simpson, J.A. (1999). Attachment theory in modern evolutionary perspective. In J. Cassidy and P.R. Shaver (eds), *Handbook of Attachment: Theory, Research, and Clinical Applications*. Guilford, New York, pp. 115–140.
Skolnick, A. (1986). Early attachment and personal relationships across the life span. In T. Balthes, D. Featherman and R. Lerner (eds), *Life Span Development and Behavior*. Erlbaum, Hillsdale, NJ, pp. 173–206.
Smith, P.B. and Pederson, D.R. (1988). Maternal sensitivity and patterns of infant–mother attachment. *Child Dev.* **59**, 1097–1101.
Solomon, J. and George, C. (eds) (1999). *Attachment Disorganization*. Guilford, New York.
Spangler, G. and Grossmann, K.E. (1993). Biobehavioral organization in securely and insecurely attached infants. *Child Dev.* **64**, 1439–1450.
Spieker, S.J. (1988). Patterns of very insecure attachment in samples of high-risk infants and toddlers. *Topics Early Childhood Spec. Educ.* **6**, 37–53.
Spieker, S.J. and Booth, C. (1988). Maternal antecedents of attachment quality. In J. Belsky and T. Nezworski (eds), *Clinical Omplications of Attachment*. Erlbaum, Hillsdale, NJ, pp. 95–135.
Spitzer, R.L., Endicott, J. and Gibbon, M. (1979). Crossing the border into borderline personality and borderline schizophrenia: the development of criteria. *Arch. Gen. Psychiat.* **36**, 17–24.
Sroufe, L.A. (1979). The coherence of individual development. *Am. Psychol.* **34**, 834–841.
Sroufe, L.A. and Waters, E. (1977). Attachment as an organizational construct. *Child Dev.* **48**, 1184–1199.
Stearns, S. (1992). *The Evolution of Life Histories*. Oxford University Press, New York.
Steele, H., Steele, M. and Fonagy, P. (1996). Associations among attachment classifications of mothers, fathers, and their infants. *Child Dev.* **67**, 541–555.
Stern, D.N. (1985). *The Interpersonal World of the Infant*. Basic Books, New York.
Stern, D.N. (1994). One way to build a clinically relevant baby. *Inf. Ment. Health J.* **15**, 36–54.
Stone, M.H. (1983). Long-term outcome in personality disorders. *Br. J. Psychiat.* **162**, 299–313.
Trivers, R. (1971). The evolution of reciprocal altruism. *Qu. Rev. Biol.* **46**, 35–57.
Trivers, R.L. (1974). Parent–offspring conflict. *Am. Zool.* **14**, 249–264.
Trivers, R. (2000). The element of a scientific theory of self-deception. In D. LeCroy and P. Moller (eds), *Evolutionary Perspectives on Human Reproductive Behavior. Ann. N. Y. Acad. Sci.* **907**, 114–131.
Tronick, E., Cohn, J. and Shea, E. (1986). The transfer of affect between mothers and infants. In B.T. Brazelton, M. Yogman (eds), *Affective Development in Infancy*. Ablex, Norwood, NJ, 11–26.
van der Kolk, B.A. (1994). The body keeps the score: memory and the evolving psychobiology of posttraumatic stress. *Harvard Rev. Psychiat.* **1**, 253–265.
van Ijzendoorn, M. (1995). Adult attachment representations, parental responsiveness, and infant attachment: a meta-analysis on the predictive validity of the Adult Attachment Interview. *Psychol. Bull.* **117**, 387–403.

van Ijzendoorn, M.H. and Sagi, A. (1999). Cross-cultural patterns of attachment: universal and contextual dimensions. In J. Cassidy and P. Shaver (eds), *Handbook of Attachment*. Guilford, New York, pp. 713–734.

van Ijzendoorn, M.H., Feldbrugge, J.T.T.M., Derks, F.C.H., DeRuiter, C. *et al*. (1997). Attachment representations of personality-disordered criminal offenders. *Am. J. Orthopsychiat*. **67**, 449–459.

Vygotsky, L.S. (1978). *Mind in society: the development of higher psychological processes*. Harvard University Press, Cambridge, MA.

Ward, M.J. and Carlson, E. (1995). Associations among adult attachment representations, maternal sensitivity, and infant–mother attachment in a sample of adolescent mothers. *Child Dev*. **66**, 69–79.

Waters, E., Wippman, J. and Sroufe, L.A. (1979). Attachment, positive affect, and competence in the peer group: two studies in construct validation. *Child Dev*. **50**, 821–829.

Winnicott, D.W. (1965). *The Maturational Process and the Facilitating Environment*. International Universities Press, New York.

Wright, R. (1994). *The Moral Animal*. Vintage, New York.

Wurmser, L. (2000). *The Power of the Inner Judge: Psychodynamic Treatment of the Severe Neuroses*. Jason Aronson, Northvale, NJ.

Youngblade, L.M. and Belsky, J. (1989). Child maltreatment, infant–parent attachment security, and dysfunctional peer relationships in toddlerhood. *Topics Early Childhood Spec. Educ*. **9**, 1–15.

Zeifman, D. and Hazan, C. (1997). Attachment: the bond in pair-bonds. In J.A. Simpson and D.T. Kenrick (eds), *Evolutionary Social Psychology*. Erlbaum, Mahwah, NJ, pp. 237–263.

18

Awareness and Theory of Mind in Dementia

SERGIO E. STARKSTEIN[1,2] AND MARÍA LAURA GARAU[1]

[1]Buenos Aires Neuropsychiatric Center, Argentina, and [2]University of Western Australia, Perth, Australia.

> I know that I exist; the question is, what is this 'I' that I know? (Descartes 1641)

Whereas for the German philosopher Fichte (1794) 'the I exists only insofar as it is conscious of itself . . .', the American philosopher William James (1950) considered the 'self . . . to consist mainly of . . . peculiar motions in the head or between the head and throat', and the British philosopher Anthony Kenny (1988) considered the self as '. . . a mythical entity . . . to allow the space which differentiates 'my self' from 'myself' to generate the illusion of a mysterious entity distinct from . . . the human being'. The above helps to illustrate the unclear ontological status of the 'self', i.e. whether it is a 'what' or a 'how'.

The concept of consciousness faces a similar conflicting state. 'Consciousness' was defined by the English philosopher John Locke (1690) as 'the perception of what passes in a man's own mind', and 'reflection' was defined as 'that notice which the mind takes of its own operations and the manner of them'. This ability was later re-named 'introspection' and considered a special way of knowing. Another English philosopher, Gilbert Ryle (1984), denied the existence of 'introspection' but admitted the ability of 'retrospection' (i.e. a capacity to recall one's states immediately after they occur). More recent philosophers accept the validity of both processes, introspection and retrospection, and stress that these abilities pertain to the individual mental state.

Within the field of neuropsychiatry, the term 'awareness' is used to refer to the processes usually subsumed by philosophers under the names of

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

'introspection' and 'retrospection'. From a mechanistic point of view, self-awareness may require the synthesis of information from countless neuronal networks involved in sensorial, memorial and emotional processing. Considering the extensive neuronal damage taking place in dementia, it is not surprising that impairment of awareness may be an early and prominent finding of this illness.

Patients with dementia not only develop a progressive decline of their intellectual capacities and motivation, but may show no awareness of their own limitations in everyday life. The term 'anosognosia' is generically used to refer to unawareness of deficits (e.g. sensory, attentional, cognitive, etc.). Demented patients also develop a progressive impairment to recognise the physical and personality characteristics of relatives and caregivers: not only do they fail to recognise physical attributes such as faces or voices, but may distort and fabulate others' intentions and actions as well. Deficits in the ability to infer other people's mental states, thoughts and feelings are usually referred to as 'theory of mind' (ToM), and may be a prominent early finding in demented patients. The main aim of the present chapter is to review recent findings on the frequency, clinical correlates and potential mechanisms of anosognosia and ToM deficits in dementia.

DEFICITS OF AWARENESS IN DEMENTIA

Awareness of one's own cognitive problems may be already disrupted in the early stages of dementia. Patients with mild memory problems, as referred by caregivers, may not report those problems spontaneously or may tend to minimise them. Some patients are brought to consult against their will, while others may show denial of illness associated with behavioural changes such as paranoid ideation, irritability and agitation. At the other extreme are patients with dementia and mild severity of cognitive deficits who are well aware of their cognitive deficits and worry about their problem, suggesting that anosognosia is a problem restricted to a subset of patients with dementia.

Reisberg, Gordon and McCarthy (1985) examined the association between anosognosia and severity of cognitive deficits, and found significantly better awareness of cognitive deficits in patients with mild cognitive decline as compared to patients with moderate to severe dementia. The authors suggested that denial of illness could be the product of a defence mechanism, 'protecting' Alzheimer's disease (AD) patients against depressive feelings.

To assess anosognosia in dementia, Migliorelli *et al.* (1995) designed the Anosognosia Questionnaire — Dementia (AQ-D), which consists of 30 questions divided into two sections. The first section assesses intellectual functioning, while the second section examines changes in interests and personality (Table 18.1). There are two forms for this questionnaire, Form A,

answered by the patient alone, and Form B (a similar questionnaire written in the third person), answered by the patient's caregiver, blind to the patient's answers in Form A. The final score is obtained by subtracting the scores on Form B from those on Form A. Thus, positive scores indicate that the caregiver rated the patient as more impaired than the patient's own evaluation (i.e. the patient was less aware of her/his cognitive and emotional deficits). Using this instrument and a comprehensive neuropsychological evaluation, Migliorelli *et al*. (1995) assessed a consecutive series of 103 patients attending a dementia clinic. Based on a healthy age-comparable population, the authors diagnosed anosognosia using a cut-off score on the AQ-D, and found that 21 of the 103 patients (20%) showed moderate to severe anosognosia. Anosognosia was significantly associated with a longer duration of illness, more severe impairments in activities of daily living, more severe deficits in long-term verbal memory, and male gender. Anosognosia was also significantly related to specific psychiatric disturbances, such as higher scores on mania and pathological laughing, but a significantly lower frequency of dysthymia (but not major depression). This finding of more severe anosognosia and less dysthymia would support the contention that dysthymia in AD is an emotional reaction to the perceived cognitive decline, but the lack of a significant negative correlation between anosognosia and major depression suggests that biological dysfunction could underlie major depression in AD (this possibility is further discussed below). The finding of a significant association between anosognosia and higher manic and pathologic laughing scores also suggests that anosognosia in dementia may be part of a wider psychiatric syndrome characterised by loss of insight, elevated mood, and abnormal release of positive emotional display.

One limitation to the study of anosognosia in AD is that most investigations diagnosed this phenomenon based on whether patients were unaware of the cognitive deficits, but whether patients with AD may also show unawareness of behavioural problems has rarely been examined. Starkstein *et al*. (1996) examined different domains of anosognosia in a study that included a consecutive series of 170 patients with AD. All the AQ-D items were entered into a principal components factor analysis, and a two-factor solution was derived: factor 1 accounted for 31% of the variance and was construed as an unawareness of cognitive deficit factor, which included unawareness of memory, temporal and spatial orientation, calculation, abstract reasoning, and praxis deficits; whereas factor 2 accounted for 7% of the variance and was construed as an unawareness of behavioural problems factor, which included selfishness, irritability, inappropriateness of emotional display and instinctive disinhibition. Thus, this study confirmed two dimensions of anosognosia in AD, namely unawareness of cognition impairments and unawareness of behavioural problems, as independent constructs. The study further demonstrated that unawareness of cognitive impairments was significantly correlated

Table 18.1. Anosognosia Questionnaire — Dementia (patient version)

A. Intellectual functions

1. Do you have problems remembering the date?
   Never Sometimes Often Always
2. Do you have problems orientating yourself in new places?
   Never Sometimes Often Always
3. Do you have problems remembering telephone calls?
   Never Sometimes Often Always
4. Do you have problems understanding conversations?
   Never Sometimes Often Always
5. Do you have problems signing your name?
   Never Sometimes Often Always
6. Do you have problems understanding what you read in the newspaper?
   Never Sometimes Often Always
7. Do you have problems keeping your personal belongings in order?
   Never Sometimes Often Always
8. Do you have problems remembering where you leave things in your house?
   Never Sometimes Often Always
9. Do you have problems writing notes or letters?
   Never Sometimes Often Always
10. Do you have problems handling money?
    Never Sometimes Often Always
11. Do you have problems orientating yourself in your neighbourhood?
    Never Sometimes Often Always
12. Do you have problems remembering appointments?
    Never Sometimes Often Always
13. Do you have problems practising your favourite hobbies?
    Never Sometimes Often Always
14. Do you have problems communicating with people?
    Never Sometimes Often Always
15. Do you have problems doing mental calculations?
    Never Sometimes Often Always
16. Do you have problems remembering things you have to buy when you go shopping?
    Never Sometimes Often Always
17. Do you have problems controlling your sphincters?
    Never Sometimes Often Always
18. Do you have problems understanding the plot of a movie?
    Never Sometimes Often Always
19. Do you have problems orientating in your house?
    Never Sometimes Often Always
20. Do you have problems doing home activities (cooking, cleaning, fixing things, etc.)?
    Never Sometimes Often Always
21. Do you have problems feeding yourself?
    Never Sometimes Often Always
22. Do you have problems keeping your chequebook, accounts, payments, etc?

B. Behaviour

23. Are you more rigid in your decisions, with less capacity to adapt to new situations?
    Never Sometimes Often Always

(*continued*)

Table 18.1. *Continued*

24.	Are you more egotistic, paying less attention to other people's needs?			
	Never	Sometimes	Often	Always
25.	Are you more irritated? Do you easily lose your temper?			
	Never	Sometimes	Often	Always
26.	Do you have crying episodes?			
	Never	Sometimes	Often	Always
27.	Do you laugh in inappropriate situations?			
	Never	Sometimes	Often	Always
28.	Are you more interested in sexual themes, talking or reading about sex?			
	Never	Sometimes	Often	Always
29.	Have you lost interest in hobbies or activities you used to like?			
	Never	Sometimes	Often	Always
30.	Do you feel more depressed?			
	Never	Sometimes	Often	Always

to deficits of verbal memory and verbal comprehension and a longer duration of illness, suggesting that cognitive deficits may account for some aspects of the anosognosia syndrome in AD. On the other hand, there were no significant correlations between unawareness of behavioural problems and cognitive impairments.

Interestingly, both types of anosognosia were significantly associated with specific behavioural problems. Whereas unawareness of cognitive deficits was significantly correlated with more severe delusions and apathy and less depression, unawareness of behavioural problems was significantly correlated with more severe disinhibition and pathological laughing. There also was a significant correlation between unawareness of cognitive deficits and depression scores, suggesting a role for depression in this type of anosognosia.

In conclusion, anosognosia is an early finding in dementia and may be present in about 20% of cross-sectional samples of patients. Demented patients may be unaware not only of their cognitive deficits but also of their behavioural changes. These two types of anosognosia are independent phenomena, have different clinical correlates and may have a different mechanism.

MECHANISM OF ANOSOGNOSIA IN DEMENTIA

The question now arising is whether behavioural problems significantly associated with anosognosia in dementia may account for this phenomenon. For instance, apathetic patients may be less reactive to both their context and their needs and emotions, and this decreased reactivity may play a part in the production of anosognosia. Delusions are false beliefs that are held despite contextual evidence to the contrary, and constitute a frequent finding among

patients with dementia. Delusions may result from dysfunction of polymodal association areas involved in the assessment of the patient's context, and anosognosia could occur with further disruption of brain areas related to self-assessment.

In Starkstein *et al.*'s (1996) study, unawareness of behavioural problems was significantly associated with symptoms of disinhibition, and could be part of a syndrome characterised by the release of inappropriate desires, beliefs and behaviours, such as grandiose ideas, irritability, hyperactivity and inappropriate emotional display. The brain mechanisms that regulate the release of behaviours could simultaneously activate brain areas that mediate awareness of activated behaviours, and dysfunction of these areas may explain the association of behavioral disinhibition and anosognosia.

Dalla Barba *et al.* (1995) assessed frontal functions, anosognosia and intrusions in patients with AD. They found that patients with anosognosia produced significantly more intrusions, which were positively and significantly correlated with anosognosia for memory deficits. The only frontal task to correlate with anosognosia was verbal fluency. They concluded that whereas anosognosia of memory deficits is necessary for intrusions to occur, frontal dysfunction is not a necessary condition for intrusions or anosognosia.

Other studies reported significant frontal lobe dysfunction underlying anosognosia in dementia. Michon *et al.* (1994) found a significant correlation between more severe anosognosia and more deficits on the Wisconsin Card Sorting Test (WCST), which evaluates frontal executive functions. Other frontal lobe-related measures, such as verbal fluency, Luria's graphic series and 'frontal behaviours' (e.g. inertia, indifference, imitation behaviours, prehension and utilisation), also showed a high correlation with anosognosia severity. Starkstein *et al.* (1997a) reported that patients with severe anosognosia performed significantly worse on the WCST and measures of procedural learning, as compared to AD patients without anosognosia. Ott *et al.* (1996) confirmed that unawareness of cognitive deficits in dementia correlated significantly with more severe deficits on tests of executive and visuospatial function. Evidence of frontal lobe dysfunction in the mechanism of anosognosia also comes from neuroimaging studies. Using single photon emission computed tomography (SPECT) to measure regional cerebral blood flow in AD, anosognosia was reported to be significantly correlated with diminished right dorsolateral and ventral frontal perfusion (Reed, Jagurt and Coulter 1993; Starkstein *et al.* 1995).

Stuss and Alexander (2000) proposed a hierarchical model for awareness at four operational levels: arousal–attention, perceptual–motor, executive mediation and self-awareness, and suggested the two highest levels to be related to frontal lobe functioning, although mediated by different circuits. Rhesus monkeys with frontal ablations showed deficits in recognising their own reflections in a mirror as images of themselves, and reacted with social

responses as though they were confronted with other monkeys (Gallup and Suárez 1991). Wheeler, Stuss and Tulving (1997) proposed that prefrontal areas, particularly on the right, provide essential organisation for episodic memory, guiding the recall of experiences and their associated emotions, and connect them with plans and expectations for the future, giving a sense of continuity of the self across time.

LONGITUDINAL EVOLUTION OF ANOSOGNOSIA IN AD

Starkstein *et al*. (1997b) examined the longitudinal evolution of anosognosia in a series of 62 AD patients who had a follow-up evaluation between 1 and 2 years after the initial evaluation. Based on the AQ-D scores, patients were divided into groups with either no, mild or severe anosognosia. A two-way ANOVA with repeated measures (group × time) showed a significant group effect — patients with severe anosognosia had significantly higher overall scores of anosognosia than the other two groups; there was a significant time effect — a significant increase in anosognosia scores during the follow-up period; and there was a significant group × time interaction — whereas patients with either no or mild anosognosia showed significant increments in anosognosia scores, patients with severe anosognosia at the initial evaluation had no further increments in anosognosia scores at follow-up. These findings suggest that, whereas anosognosia may be heterogeneously distributed in the early stages of dementia, most patients show a progressive deficits of awareness with the progression of cognitive decline. Thus, a specific level of cognitive deficits may not be necessary for the production of anosognosia, since even patients with very mild cognitive deficits may show this phenomenon, although a detailed cognitive evaluation could show a significant association between anosognosia and specific cognitive deficits. On the other hand, a certain amount of cognitive deficits may be sufficient to produce anosognosia, since most patients without or with mild anosognosia at initial evaluation will eventually develop significant anosognosia with the progression of illness.

The question then arises as to whether the anosognosia present at the initial stages of the illness, and the slowly progressing anosognosia developing along the progression of the illness, have a common mechanism. Several studies showed a significant association between anosognosia in AD and perfusion deficits in specific brain areas. Starkstein *et al*. (1995) reported that AD patients with severe anosognosia had a significantly lower right frontal perfusion than AD patients without anosognosia matched for age, duration of illness and cognitive impairments, and speculated that AD patients with relatively more severe right frontal hypoperfusion may develop anosognosia early in the disease; whereas AD patients with less severe or no anosognosia during the initial stages of the illness may develop anosognosia in later stages, with further disruption of right frontal lobe functions.

THEORY OF MIND AND PRAGMATIC ABILITIES IN DEMENTIA

'Theory of mind' (ToM) is defined as the capacity to attribute mental states to oneself and to others, and to interpret behaviours in terms of mental states. Normal social behaviour could result from a specific interaction between cognitive, behavioural and emotional abilities, including the understanding of thoughts and feelings of others. Walston, Blennerhassett and Charlton (2000) suggested that errors in the inference of another person's mental state may produce delusions, and these errors could be most frequent when emotions are inappropriate to a specific social situation. Given the progressive cognitive decline that characterises dementia, AD patients could show deficits in their ability to make inferences about others' representational states, and to predict and select behaviours accordingly. Thus, the study of ToM could provide some insight into the abilities of social cognition in dementia, and may help to determine whether behavioural abnormalities frequently found in AD, such as apathy, irritability and paranoid thinking, are related to deficits in pragmatic abilities.

García-Cuerva *et al.* (2001) assessed first-order and second-order ToM tasks in a series of 34 patients with AD and a group of age-comparable healthy controls. To minimise the influence of cognitive deficits in the assessment of ToM, only AD patients with MMSE scores $\geqslant 17$ points were included into the study. One of the main findings was that only 35% of the AD patients but all 10 healthy controls passed the second-order ToM task ($p < 0.01$). AD patients who did not pass the second-order ToM task (i.e. those without ToM) had significantly more severe cognitive deficits than AD patients that passed the task, suggesting that the difference in ToM performance may have resulted from relatively more severe cognitive deficits in the former group. However, AD patients without ToM also had significantly lower scores than the AD group with ToM on relatively easier first-order ToM tasks, suggesting that generic cognitive dysfunction may not fully account for deficits in second-order ToM in AD. Another relevant finding was the lack of significant differences on behavioural variables between AD patients with or without ToM, suggesting that mentalising deficits may not explain the behavioural problems in AD.

Garcia-Cuerva *et al.* (2001) examined deficits in language pragmatics in a series of 39 patients with AD, using tasks assessing comprehension of indirect requests and conversational implications. The main finding was that AD patients (with or without ToM) had a significantly worse performance on tests of indirect requests and conversational implications than age-comparable healthy controls (Figures 18.1, 18.2). Interestingly, AD patients without ToM has significantly more deficits on conversational implications, as compared to AD patients with preserved ToM, suggesting a significant association between pragmatic abilities and ToM.

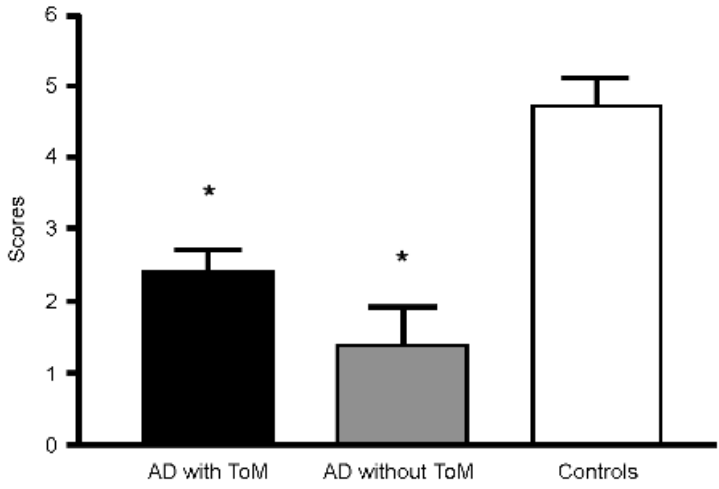


Figure 18.1. AD patients with or without theory of mind (ToM) showed significantly more severe deficits on a task assessing indirect requests than age-comparable healthy controls

Gregory *et al.* (2002) assessed ToM, executive functions and general neuropsychological ability in 19 patients with the frontal variant of frontotemporal dementia (fv-FTD) and 12 patients with AD. They found that the degree of impairment of ToM in fv-FTD patients was related to the level of neurobehavioural disturbance, and the only significant correlation between ToM and frontal executive functions was with the number of perseverative errors on the WCST. None of the other executive functions, semantic memory or general intellectual measures were significantly correlated with ToM performance. Both AD and fv-FTD groups showed significantly more severe deficits on second-order ToM tests as compared to healthy controls, but no significant between-group differences were found on first-order ToM tasks. There was also a significant association between deficits on ToM tasks and severity of ventromedial frontal cortex atrophy in fv-FTD patients.

Patients with dementia may not be aware of the full meaning of social situations, which could result in behavioural problems, such as paranoid thinking and aggressive outbursts. In a series of 25 AD patients and 20 age-comparable healthy controls, Torralva *et al.* (2000) measured social cognition using a social dilemma and, based on the responses to this task, calculated a stage of 'moral development'. One of the main findings of the study was that AD patients showed a significantly lower overall score on the moral judgement task as compared to healthy controls. Moreover, whereas all healthy controls were classified in high levels of moral judgement, no AD patient achieved those levels. There were no significant correlations between scores of moral

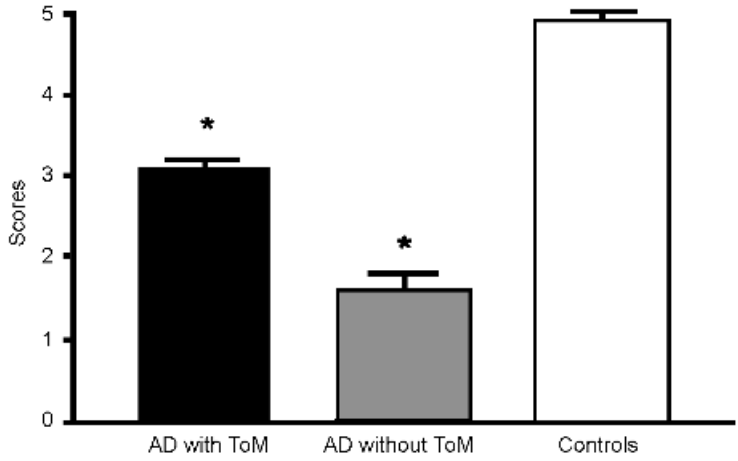


Figure 18.2. Whereas AD patients with or without ToM showed significantly more severe deficits on a task assessing conversational implications than age-comparable healthy controls, AD patients without ToM had a significantly worse performance than AD patients with ToM

judgement and scores on psychiatric scales; but there were significant correlations between scores of moral judgement and scores on a test of abstract reasoning (the Raven's Progressive Matrices). Taken together, these findings suggest that whereas deficits in moral judgement are highly prevalent in AD, they do not relate to behavioural disorders in dementia and may depend on deficits of more basic intellectual functions.

One limitation to the study of behavioural disorders in dementia is the scarcity of a cognitive measure correlating with these disorders. Some tasks designed for this purpose are quite demanding and difficult to adapt to dementia patients. Bechara *et al.* (1994) designed a neuropsychological task that simulates personal real-life decision-making related to punishment and rewards, which is important for risk-evaluation behaviour. This task requires subjects to select cards from four different decks, and each card selection is associated with gains or losses. The aim of the task is to maximise gains, which patients may achieve by avoiding decks yielding high gains but higher losses, and selecting from more conservative decks. Torralva *et al.* (2000) assessed the Bechara's Card Test in a series of 25 patients with AD and 20 age-comparable normal controls, and AD patients also underwent a comprehensive psychiatric and neuropsychological evaluation. The main finding of the study was that AD patients obtained significantly lower gains on the Bechara's Card Test than the healthy control group. The AD group selected more cards from the 'high-risk' decks than did controls, but both groups had a similar pattern of earnings

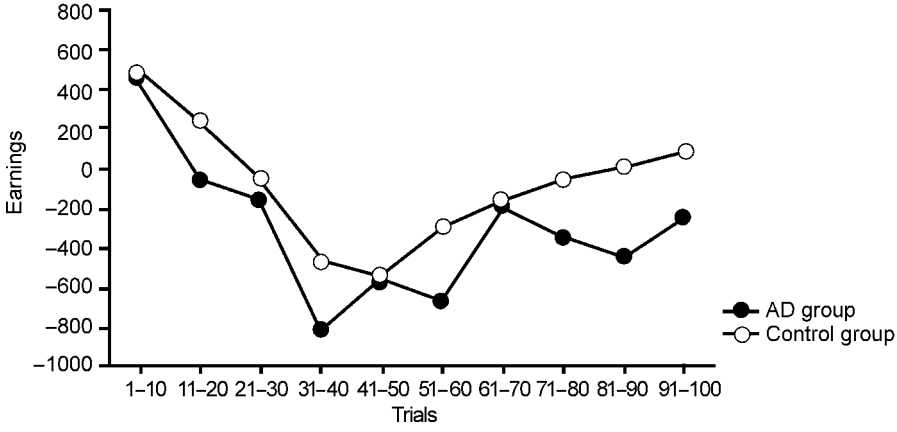


Figure 18.3. Card game test; earnings across trials. AD patients obtained significantly less earnings than the control group, but this lower performance was stable across the whole trial

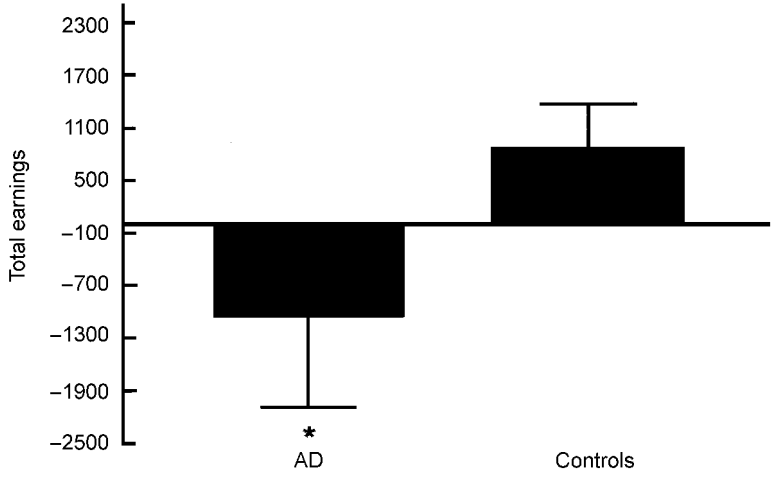


Figure 18.4. AD patients obtained significantly lower total earnings on Bechara's card test as compared to the age-comparable normal control group ($^{*}F(1.36) = 42.6$, $p < 0.0001$)

along the test (Figures 18.3, 18.4). A regression analysis with earnings on the Card Test as the dependent variable and scores on psychiatric scales as independent variables showed no significant overall effect, suggesting that deficits in real-life decision-making may not underlie the presence of frequent behavioural problems in dementia. On the other hand, there were significant correlations between total earnings and deficits in both verbal and visuospatial

memory tests (i.e. the lower the memory scores, the lower the earnings on the Card Test), suggesting that cognitive deficits may account for some of the impairments on the Card Test among demented individuals. Rahman *et al.* (1999) assessed the Bechara Card Test in eight patients with frontotemporal dementia (FTD) and eight age-comparable healthy controls. They found FTD patients to bet a much higher proportion of their accumulated reward as compared to the healthy control group, but there were no significant between-group differences on the choice of the most likely outcome.

CONCLUSIONS

Unawareness of cognitive and behavioural changes is a frequent finding in patients with dementia. Anosognosia in dementia may be part of a psychiatric syndrome, including elevated mood, irritability and the release of positive emotional display. Some patients may show anosognosia already in the early stages of dementia, whereas most patients show increasing anosognosia with the progression of the illness. 'Early' anosognosia may be related to relatively more severe right frontal hypoperfusion in the initial stages of the illness, whereas anosognosia in later stages may be related to further right frontal dysfunction. Patients with dementia have significant impairments on tasks of moral judgement and risk-taking abilities, but these deficits are not related to behavioural abnormalities frequently reported in the disease. Patients with dementia also showed significant deficits in first- and second-order ToM tasks, as well as on tasks assessing verbal pragmatic abilities. Deficits on these 'social cognition' tests have a significantly stronger association with deficits on specific cognitive domains, as compared to behavioural abnormalities.

REFERENCES

Bechara, A., Damasio, A.R., Damasio, H. and Anderson, S.W. (1994) Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* **50**, 7–15.

Dalla Barba, G., Parlato, V., Iavarone, A. and Boller, F. (1995). Anosognosia, intrusions and 'frontal' functions in Alzheimer's disease and depression. *Neuropsychologia* **33**, 247–259.

Descartes, R. (1641). Quoted in *The Philosophical Writings of Descartes*, vol. 1 (I. Cottingham *et al.*, trans.). Cambridge University Press, Cambridge.

Gallup, G. Jr and Suárez, S. (1991). Social responses to mirrors in rhesus monkeys (*Macaca mulata*): effects of temporary mirror removal. *J. Comp. Psychol.* **105**, 376–379.

García Cuerva, A., Sabe, L., Kuzis, G., Tiberti, C. *et al.* (2001). 'Theory of mind' and pragmatic abilities in dementia. *Neuropsychiat. Neuropsychol. Behav. Neurol.* **14**, 153–158.

Gregory, C., Lough, S., Stone, V., Erzinclioglu, S. *et al.* (2002). Theory of mind in patients with frontal variant frontotemporal dementia and Alzheimer's disease: theoretical and practical implications. *Brain* **125**, 752–764.

James, W. (1950). *The Principles of Psychology*. Denver, New York.

Kenny, A. (1988). *The Self*. Marguette University Press, New York.

Locke, J. (1690). *An Essay Concerning Human Understanding*. Oxford University Press, Oxford.

Michon, A., Deweer, B., Pillon, B., Agid, Y. and Dubois, B. (1994). Relation of anosognosia to frontal lobe dysfunction in Alzheimer's disease. *J. Neurol. Neurosurg. Psychiat.* **57**, 805–809.

Migliorelli, R., Tesón, A., Sabe, L., Petracca, G. *et al.* (1995). Anosognosia in Alzheimer's disease: a study of associated factors. *J. Neuropsychiat. Clin. Neurosci.* **7**, 338–344.

Ott, B., Lafleche, G., Whelihan, W., Buongiorno, G. *et al.* (1996). Impaired awareness of deficits in Alzheimer's disease. *Alzheim. Dis. Assoc. Disord.* **10**, 68–76.

Rahman, S., Sahakian, B., Hodges, J., Rogers, R. and Robbins, T. (1999). Specific cognitive deficits in mild frontal variant frontotemporal dementia. *Brain* **122**, 1469–1493.

Reed, B., Jagust, W. and Coulter, L. (1993). Anosognosia in Alzheimer's disease: relationships to depression, cognitive function and cerebral perfusion. *J. Clin. Exp. Neuropsychol.* **15**, 231–244.

Reisberg, B., Gordon, B. and McCarthy, M. (1985). Insight and denial accompanying progressive cognitive decline in normal aging and Alzheimer's disease. In B. Stanley (ed.), *Geriatric Psychiatry: Ethical and Legal Issues*. American Psychiatric Press, Washington, DC.

Ryle, G. (1984). *The Concept of Mind*. Cambridge University Press, Cambridge.

Starkstein, S., Vázquez, S., Migliorelli, R., Tesón, A. *et al.* (1995). A single-photon emission computed tomographic study of anosognosia in Alzheimer's disease. *Arch. Neurol.* **52**, 415–420.

Starkstein, S., Sabe, L., Chemerinski, E., Jasón, L. and Leiguarda, R. (1996). Two domains of anosognosia in Alzheimer's disease. *J. Neurol. Neurosurg. Psychiatry* **61**, 485–490.

Starkstein, S., Sabe, L., Cuerva, A., Kuzis, G. and Leiguarda, R. (1997a). Anosognosia and procedural learning in Alzheimer's disease. *Neuropsychiat. Neuropsychol. Behav. Neurol.* **10**, 96–101.

Starkstein, S., Chemerinski, E., Sabe, L., Kuzis, G. *et al.* (1997b). Prospective longitudinal study of depression and anosognosia in Alzheimer's disease. *Br. J. Psychiat.* **171**, 47–52.

Stuss, D. and Alexander, M. (2000). Executive functions and the frontal lobes: a conceptual view. *Psychol. Res.* **63**, 289–298.

Torralva, T., Dorrego, F., Sabe, L., Chemerinski, E. and Starkstein, S. (2000). Impairments of social cognition and decision making in Alzheimer's disease. *Int. Psychogeriat.* **12**, 359–368.

Walston, F., Blennerhassett, R. and Charlton, B. (2000). 'Theory of mind', persecutory delusions and the somatic marker mechanism. *Cogn. Neuropsychiat.* **5**, 161–174.

Wheeler, M., Stuss, D. and Tulving, E. (1997). Towards a theory of episodic memory: the frontal lobes and autonoetic consciousness. *Psychol. Bull.* **121**, 331–354.

19

Postscript

MARTIN BRÜNE[1], HEDDA RIBBERT[1] AND WULF SCHIEFENHÖVEL[2]

[1]Department of Psychiatry and Psychotherapy, University of Bochum and
[2]Human Ethology Group, Max-Planck-Society Andechs, Germany

When editing the contributions to this volume, we kept the evolutionary paradigm typical for the papers as the main guiding thread through the book, with the following rationale. Since biological anthropology, human ethology, psychology, psychiatry and the various branches of the clinical neurosciences rely on such very different conceptualisations in their effort to illuminate what the human mind is all about, we felt the only consistent way to facilitate an interdisciplinary exchange of expertise about the history, functioning and dysfunctioning of the human brain had to be essentially evolutionary. In other words, if we want to make progress in human psychology and psychopathology, we must inevitably delve into the evolutionary past of our species, i.e. ask questions such as why the human brain evolved in the particular way it actually did.

The title, *The Social Brain: Evolution and Pathology*, is admittedly provocative in some respects. Yet, we are convinced that the 'social brain hypothesis' put forward by Leslie Brothers in 1990 provides a comprehensive framework for all the disciplines mentioned above, because for us as scientists whose work so critically depends on behavioural observation of human beings — be it in field studies, laboratory investigations or work in clinical settings — it is obvious that the main problem for patients and psychiatrically healthy persons is to manage coping with their social environment. On the other hand, we are aware of the problem that the social environment of our ancestral species was most likely not the only evolutionary force that ultimately led to what we call 'intelligence'. In compiling this volume, due to reasons of limited space and our wish not to lose sight of the evolutionary scenario, we

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

therefore rather neglected other important aspects of human brain evolution. Humans, like most other primate species, however, are essentially social beings, so we found it justifiable to focus on social issues.

We do not claim to provide a complete overview of all topics relating to 'sociality'. Language, for example, is probably the most outstanding human capacity that sets human beings apart from the rest of the animal world, and language certainly plays an important role in communicating purposes, beliefs, feelings and so on to our conspecifics, but sometimes it is also used to conceal our 'real' intentions. In order to do so, one not only needs a 'theory of self and other's mind' — the scientific exploration of which is at the core of this book — but also the ability of planning ahead, of remembering past situations or, as Suddendorf and Corballis (1997) put it, to be capable of mentally travelling in time, which probably necessitated the evolutionary emergence of an episodic or autobiographic memory. Moreover, we wonder whether human self-consciousness and eventually moral systems are co-evolutionary products of selection pressures from our social environment. These important issues are only marginally addressed in the present book.

Instead, we took a more narrow perspective by pinpointing 'social intelligence'. The main reasons for this are that there are good data in this research area, offering a cross-species comparison of human beings with non-human primates, with whom we share a common ancestor, the cross-cultural evidence for the psychic unity of our species, and also empirical data on psychopathological conditions which allow at least a glimpse of what happens when a distinct psychological mechanism 'has gone awry', such as the capacity to infer what others are thinking, intending or believing.

In addition, our understanding of the brain circuits involved in these functions has recently been fuelled by imaging studies of the living brain. Thus, we get quite a clear, although not perfect, picture of essential aspects of what causes human beings to be 'hypersocial' (the term created by Michael Tomasello). Indeed, humans seem so specialised in inferring other persons' intentions and dispositions that we assume purposes in situations where virtually none can be found, and even ridiculously suspect intentions in non-living objects, e.g. assuming that our computer is on strike.

Regardless of whether the social or the ecological environment was more important in our evolutionary history, it is a matter of fact that humans grew large brains, which have costs and benefits. We may safely assume, for instance, that a better processing of sensory information may be an immediate benefit of a large brain, but other beneficial steps are delayed during human ontogeny, such as the storage and use of learned material, enabling an organism to respond more flexibly to environmental stimuli. The costly side of the coin, then, is not only the large amount of energy necessary to grow a large brain but also the time to mature flexible brain functions. These constraints are particularly true for the human brain and the acquisition of social intelligence.

Thus, the ontogenetic maturation of a theory of mind 'module' in order to understand, for example, that other persons may hold false beliefs — which emerges around the age of 4 — is not sufficient, as the child must gain social competence in when and how to apply his/her knowledge, i.e. a set of so-called 'procedural rules' (Schmitt and Grammer 1997).

If we take the 'social brain hypothesis' seriously by acknowledging that the evolution of social intelligence in humans was vital in terms of survival and reproduction — the key components of all earthly evolution — and that the benefits of the social brain must have exceeded its enormous costs in both energetic and functional respects, then we can imagine how sensitive the human brain must be to any kind of disturbance, be it genetic in origin, infectious in the prenatal phase, traumatic, or psychologically triggered by emotional neglect or lack of a stimulating environment.

We believe that these considerations create new prospects for research. 'Theory of mind', for instance, is probably not a homogeneous concept regarding its representation in the brain or with respect to its hierarchically functional organisation, neither is the definition of the term and its synonyms satisfactory. On the one hand, 'theory of mind' or 'enphronesis' has to be distinguished from 'empathy', the former being the capacity to represent intentions and beliefs, the latter being the ability to feel what others are feeling. On the other hand, however, the separation of 'emotion' from 'cognition' in psychology and the neurosciences, as if they were independent categories, has rightly been criticised, because they undoubtedly co-evolved and are closely linked in functional terms. There is good empirical evidence of profound deficits in acquiring emotional and theory of mind skills in autistic spectrum disorders and, perhaps more subtly, also in schizophrenia, but we know little about the critical causes of why social cognitive impairments emerge in personality disorders. It is likely that for some psychiatric disorders the genetic predisposition is crucial, but in others it may rather be the social environment, particularly during early infancy and childhood, which leads to a deficient learning of procedural rules; hence, antisocial behaviour and other behavioural disturbances may develop, partly due to an impaired mentalising capacity. Therefore, it could be worthwhile to study the relationship of the ontogeny of enphronesis and empathy with biographical data, rearing conditions, etc. With regard to therapy, it could be useful to include the concept of the 'social brain' in educational programs, such as social skills training for patients with psychotic disorders, in ways that specifically address their abilities to 'read' the emotions and the intentions of other persons. Moreover, to date, neuropsychology and neuropsychiatry have neglected conveying their findings to cognitive-behavioural therapy of non-psychotic disorders. In an experimental design, however, the concept of 'theory of mind' may usefully be integrated at a meta-level, where patients with deficits of theory of mind performance (in the procedural sense) receive information about the nature of their difficulties, in

ways similar to those well established in the cognitive-behavioural therapy of anxiety disorders and obsessive-compulsive disorder.

We find it promising to integrate enphronesis as a key process of social interaction more systematically into psychotherapeutic settings (it is noteworthy that this is actually partially done, particularly in modern psychodynamic therapy of personality disorders), proposing that patients' resources in this domain can be stimulated and encouraged in many cases. Also, in terms of the prevention of mental illness in general, we may be better off in the long term if we create environments that meet the needs of our children to learn social exchange, including tolerance and empathy, which we believe all human primates deserve.

REFERENCES

Brothers, L. (1990). The social brain: a project for integrating primate behavior and neurophysiology in a new domain. *Concepts Neurosci.* **1**, 27–51.

Schmitt, A. and Grammer, K. (1997). Social intelligence and success: don't be too clever in order to be smart. In A. Whiten and R.W. Byrne (eds), *Machiavellian Intelligence II. Extensions and Evaluations*. Cambridge University Press, Cambridge, pp. 86–111.

Suddendorf, T. and Corballis, M.C. (1997). Mental time travel and the evolution of the human kind. *Genet. Soc. Gen. Psychol. Monogr.* **123**, 133–167.

Index

Note: Page numbers in *italics* refer to figures; page numbers in **bold** refer to tables; 'n' after a page number signifies a footnote.

The Social Brain: Evolution and Pathology. Edited by M. Brüne, H. Ribbert and W. Schiefenhövel.
 ISBN 0 470 84960 6

Index compiled by Lewis N. Derrick