

The Structure and Development of Self-Consciousness

Advances in Consciousness Research

Edited by Dan Zahavi,
Thor Grünbaum and
Josef Parnas



The Structure and Development of Self-Consciousness

Advances in Consciousness Research

Advances in Consciousness Research provides a forum for scholars from different scientific disciplines and fields of knowledge who study consciousness in its multifaceted aspects. Thus the Series will include (but not be limited to) the various areas of cognitive science, including cognitive psychology, linguistics, brain science and philosophy. The orientation of the Series is toward developing new interdisciplinary and integrative approaches for the investigation, description and theory of consciousness, as well as the practical consequences of this research for the individual and society.

Series B: Research in progress. Experimental, descriptive and clinical research in consciousness.

Editor

Maxim I. Stamenov
Bulgarian Academy of Sciences

Editorial Board

David Chalmers
University of Arizona

Gordon G. Globus
University of California at Irvine

Ray Jackendoff
Brandeis University

Christof Koch
California Institute of Technology

Stephen Kosslyn
Harvard University

Earl Mac Cormac
Duke University

Steven Macknik
Barrow Neurological Institute

George Mandler
University of California at San Diego

Susana Martinez-Conde
Barrow Neurological Institute

John R. Searle
University of California at Berkeley

Petra Stoerig
Universität Düsseldorf

Volume 59

The Structure and Development of Self-Consciousness:
Interdisciplinary perspectives

Edited by Dan Zahavi, Thor Grünbaum and Josef Parnas

The Structure and Development of Self-Consciousness

Interdisciplinary perspectives

Edited by

Dan Zahavi

Thor Grünbaum

Josef Parnas

Danish National Research Foundation:
Center for Subjectivity Research

John Benjamins Publishing Company
Amsterdam/Philadelphia



™ The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

The structure and development of self-consciousness : interdisciplinary perspectives / edited by Dan Zahavi, Thor Grünbaum, Josef Parnas.

p. cm. (Advances in Consciousness Research, ISSN 1381-589X ; v. 59)

Includes bibliographical references and index.

I. Self-preception--Congresses. I. Zahavi, Dan. II. Grünbaum, Thor & Parnas, Josef. III. Series.

BF697.5.S43 S78 2004

153.7'5-dc22

2004055069

ISBN 90 272 5195 9 (Eur.) / 1 58811 571 2 (US) (Hb; alk. paper)

© 2004 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Table of contents

Acknowledgments	VII
The ambiguity of self-consciousness: A preface <i>Thor Grünbaum and Dan Zahavi</i>	IX
CHAPTER 1	
The emergence of self-awareness as <i>co-awareness</i> in early child development <i>Philippe Rochat</i>	1
CHAPTER 2	
Threesome intersubjectivity in infancy: A contribution to the development of self-awareness <i>Elisabeth Fivaz-Depeursinge, Nicolas Favez, and France Frascarolo</i>	21
CHAPTER 3	
The embodied self-awareness of the infant: A challenge to the theory-theory of mind? <i>Dan Zahavi</i>	35
CHAPTER 4	
From self-recognition to self-consciousness <i>Marc Jeannerod</i>	65
CHAPTER 5	
Agency, ownership, and alien control in schizophrenia <i>Shaun Gallagher</i>	89

CHAPTER 6

Tetraplegia and self-consciousness

105

Jonathan Cole

CHAPTER 7

Self and identity

123

Arne Grøn

Index

157

Acknowledgments

In May 2003, a conference entitled *Interdisciplinary perspectives on self-consciousness* was organized by Josef Parnas, Dan Zahavi, and Pierre Bovet from the Danish National Research Foundation: *Center for Subjectivity Research* at the University of Copenhagen. The aim of the conference was to discuss recent research on self-consciousness, and to contribute to a better integration of the different empirical and conceptual perspectives.

Some of the papers presented on that occasion can be found in this volume. The conference was financially supported by the Graduate School of Neuroscience and by the Danish National Research Foundation. The Center's secretary Pia Kirkemann Hansen was of invaluable assistance in the organization of the conference. Deborah Licht did an inestimable job in proof-reading the manuscripts. Finally, thanks are due to Maxim Stamenov, the editor of *Advances in Consciousness Research*, and Bertie Kaal, from John Benjamins Publishing Company, for their help with the preparation of this volume.

The ambiguity of self-consciousness

A preface

Thor Grünbaum and Dan Zahavi

University of Copenhagen

Self-consciousness is a topic of considerable importance to a variety of empirical and theoretical disciplines, such as developmental and social psychology, cognitive neuroscience, psychiatry, and philosophy. The fact that self-consciousness is being studied in many contexts and disciplines is gradually increasing our understanding of the phenomenon, but it is also complicating the situation. Thus, it would be something of an exaggeration to claim that the concept is unequivocal and that there is widespread consensus about what exactly it means to be self-conscious. Quite to the contrary, it is a simple fact that the concept connotes different things in different disciplines – sometimes radically different things. To put it differently, the psychological, neuroscientific, and philosophical literature is overflowing with competing, conflicting, and complementary definitions of self-consciousness. Let us take a quick look at some of the main candidates.

In philosophy, many have sought to link self-consciousness to the ability to think ‘I’-thoughts. A recent defense of such an approach has been put forth by Baker, who has argued that all sentient beings are subjects of experience, they all have perspectival attitudes, and they all experience the world from their own egocentric perspectives. In doing so, they show themselves to be in possession of what Baker calls *weak first-person phenomena* (Baker 2000: 60, 67). However, merely having a subjective point of view is not enough for having self-consciousness. In order to possess self-consciousness, in order to have what Baker calls *strong first-person phenomena*, one must be able to think of oneself *as* oneself. It is not enough to have desires and beliefs, it is not enough to have a perspectival attitude, it is not enough to be able to distinguish between self and non-self, one must also be able to *conceptualize* this distinction.

This is why Baker has argued that self-consciousness presupposes the possession of a first-person *concept*. One is only self-conscious the moment one can *conceive* of oneself *as* oneself, and one must have the linguistic ability to use the first-person pronoun to refer to oneself (Baker 2000:67–68). Given this definition, self-consciousness is consequently taken to be something that emerges in the course of a developmental process, and that depends upon the eventual acquisition of concepts and language.

If we shift terrain and move into social psychology, we will frequently encounter the claim, famously defended by Mead, that self-consciousness is a matter of becoming an object to oneself in virtue of one's social relations to others, i.e., that self-consciousness is constituted by adopting the perspective of the other towards oneself. According to this account, self-consciousness is *per se* a social phenomenon. It is not something you can acquire on your own. As Mead writes:

Consciousness, as frequently used, simply has reference to the field of experience, but self-consciousness refers to the ability to call out in ourselves a set of definite responses which belong to the others of the group. Consciousness and self-consciousness are not on the same level. A man alone has, fortunately or unfortunately, access to his own toothache, but that is not what we mean by self-consciousness. (Mead 1962: 163)

Within developmental psychology, the so-called mirror-recognition task has occasionally been heralded as the decisive test for self-consciousness. Thus, it has been argued that self-consciousness is only present from the moment when the child (at around eighteen months) is capable of recognizing his/her own mirror-image. Some, however, have raised the stakes even further and have argued that self-consciousness presupposes a theory of mind, and that one can consequently test the presence of self-consciousness by using classical theory of mind tasks, such as the false-belief task or the appearance-reality task. According to this proposal, children would only gain self-consciousness, in the proper sense of the word, around the age of four.

There are kernels of truth in all these definitions, in the sense that they all capture different important aspects of the phenomenon. However, all of them also claim that self-consciousness is primarily to be understood as an explicit way of relating to oneself, by way of concepts, symbols, images, or theories. Contrary to this approach, most of the contributors to the present volume argue for the existence of a minimal and implicit form of self-consciousness. It is in turn argued that empirical findings in developmental psychology, phenomenological analyses of embodiment, or studies of pathological self-

experiences point to the existence of a type of self-consciousness that does not require any explicit “I”-thought or self-observation, but is more adequately described as a pre-reflective, embodied form of self-familiarity.

Self-consciousness is an ambiguous concept, but the diversity of available definitions should not be seen as a mere expression of a comprehensive theoretical confusion (nor for that matter, as an argument for the non-existence of the phenomenon), rather it should be appreciated as a manifestation of the fact that self-consciousness comes in many different forms and degrees. Self-consciousness is a complex multifaceted phenomenon that calls for an integration of different complementary interdisciplinary perspectives. This is amply demonstrated by the contributions in the present volume.

In the first chapter, Philippe Rochat presents recent research supporting two general ideas. The first idea is that infants from birth manifest an implicit sense of themselves. This is contrary to what has been assumed by early theorists of child development, e.g., by Piaget or Mahler, who claimed that the starting point of children’s development is an initial state of “undifferentiation” between the infants and their environment. The second idea is that from this early sense of self, infants rapidly develop a sense of themselves as perceived by others. Reviewing empirical research, Rochat delineates the stages of a major developmental transition from the minimal form of self-awareness manifested at birth to the more complex awareness of the self as perceived by others, which normally starts to appear by the end of the child’s second year. According to Rochat, this development contains five stages, where each stage is a condition for and precursor of the following stages. Rochat demonstrates how the ability in children to evaluate themselves as others would rests on a more fundamental ability to perceive one’s own body as a differentiated entity among other environmental entities, as well as on an early ability to engage in social interaction (e.g., “the social smile”).

In Chapter 2, Elisabeth Fivaz-Depeursinge, Nicolas Favez, and France Frascarolo also see the origins of self-awareness as being closely linked with intersubjectivity. In their contribution, they present new empirical evidence in support of the idea that intersubjectivity is an important factor for the infant’s development of a sense of self already at a very early stage. It is argued that infants appear to have an innate or very early ability to share the feelings and mind-states of others. The evidence they present shows that young infants are able to engage not only with one person at a time, but with two, in this case their fathers and mothers. For example, a three-month-old infant who is sharing pleasure or interest with one parent might spontaneously turn to

the other one to share her affect with him or her too. The authors argue that such three-way interactions multiply the opportunities for differentiating the self from others, relative to the dyadic interactions that have traditionally been studied in developmental research. In fact, this process of triangulation and its corollary, threesome intersubjectivity, can be seen as constituting yet another argument that infants are in possession of a form of bodily self-awareness well before the acquisition of a theory of mind.

Dan Zahavi's focus in Chapter 3 is on the validity of the theory-theory of mind. Is it true that the experience of minded beings (be it oneself or others) requires a theory of mind? Is it true that self-awareness and intersubjectivity are theoretical, inferential, and quasi-scientific in nature? Is it true that mental states are unobservable and theoretically postulated explanatory devices that are introduced in order to help us predict and explain behavioral data? Zahavi discusses some recent empirical findings concerning infantile experiences of self and other that challenge the view that theoretical knowledge constitutes the core of what we call upon when we understand ourselves and others. However, Zahavi also points to certain limitations in the way in which developmental psychologists have been conceiving of the relation between bodily self-experience and object-consciousness, and he ultimately argues that the empirical challenge to the theory-theory of mind can be strengthened if it joins forces with some of the theoretical considerations found in phenomenology.

In Chapter 4, Marc Jeannerod sets out to investigate how we distinguish ourselves from other people. His focus is primarily on the sensory-motor mechanisms that allow us to recognize our bodies and our actions as our own. Central to this investigation are different experiments that reveal the role of sensory cues for the recognition of one's own body and actions. Jeannerod suggests that self-recognition is largely dependent on the recognition of one's own actions, and that the ability to recognize oneself as the agent of a behavior – the sense of agency – is crucial for the development of the self. Thus, for Jeannerod, the distinction between self-generated actions and actions produced by other agents, as well as the corresponding ability to attribute an action to its agent, are key functions of self-consciousness.

In Chapter 5, Shaun Gallagher likewise focuses on the relation between self-consciousness and motor experience. Gallagher argues that the distinction between the *sense of ownership* and the *sense of agency* is highly pertinent both when it comes to an elucidation of important facets of self-consciousness, but also regarding our understanding of the schizophrenic symptoms of thought insertion and delusions of control. In his contribution, Gallagher discusses the account offered by Graham and Stephens (1994) and he argues that they

are mistaken in locating the distinction between sense of ownership and sense of agency at a reflective, higher-order level of self-consciousness. Graham and Stephens provide a top-down account of the schizophrenic symptoms, which conflicts with clinical reports and experimental findings that, on the contrary, suggest that such symptoms originate at a “lower” or first-order level of self-consciousness. On the basis of both phenomenological and neuroscientific evidence, Gallagher then outlines a competing bottom-up analysis. He suggests that the disturbances primarily originate and manifest themselves at basic neurological and first-order phenomenal levels. It is the disturbances at these levels that then motivate the problems at the higher level of cognitive reflection.

In Chapter 6, Jonathan Cole examines the relation between an embodied form of self-awareness and more elaborate forms of consciousness of self (e.g., self-esteem and self-reflection). But in contrast to Rochat and Zahavi, Cole’s focus is on pathology. He discusses the drastic transformations of bodily experience that follows a complete high spinal cord injury, where people suddenly – and for the rest of their lives – become paralyzed and insentient from the neck down. While this disturbance of normal embodiment can be devastating, Cole shows how many people soon begin to explore their new mode of existence and how they are able to reach some accommodation with it. It is Cole’s central claim that tetraplegics, in their struggle to reclaim their lives, reveal something of how self-esteem, self-regard, and a consciousness of self are dependent on the body. Thus, clues to the way in which we take our normal embodiment for granted may be found in the process through which those with tetraplegia learn to care for themselves, navigate through the physical world, and maintain and develop personal and professional relationships.

In the final chapter, Arne Grøn investigates the problem of self-identity. He thereby introduces issues that go beyond the more minimal or formal definitions of self and self-consciousness discussed in many of the earlier chapters. Grøn distinguishes two approaches to the problem of personal identity. According to the first approach, the task is to account for the identity of oneself in terms of criteria, be they physical, psychological, or mixed. According to the second approach, which is the one defended by Grøn, to be a self is basically to have self-identity as a problem, or to put it differently, the problem of self-identity is part of what it is to be a self. The overall aim of the chapter is to develop this second approach by defining the self as a self-relation. The central claim is that as a self I am related to myself, but this self-relation is only possible in so far as I am related to others and to a more or less shared world. In the first step, Grøn deals with the issue of identity and otherness in terms of selfhood and alterity. In the second step, he focuses on temporality as an intrinsic

feature of selfhood. In the third and crucial step, he investigates the normative dimension of self and identity, and argues that the normative notion of self and identity presupposes a non-normative one.

References

- Baker, L. R. (2000). *Persons and Bodies*. Cambridge: Cambridge University Press.
- Mead, G. H. (1962). *Mind, Self and Society. From the standpoint of a social behaviorist*. Chicago: University of Chicago Press.

CHAPTER 1

The emergence of self-awareness as *co-awareness* in early child development

Philippe Rochat
Emory University

This chapter presents some recent observations supporting two general ideas. The first idea is that, contrary to what has been assumed by early theorists of child development, from the first few weeks of life, infants manifest an implicit sense of themselves. They perceive their own bodies as differentiated, situated, and agent entities among other non-self entities in the environment, namely physical objects and people. The second idea is that from this early sense of self, infants develop rapidly a sense of themselves *as perceived by others*. They develop what I call “co-awareness”, or the *awareness of self in relation to and through the eyes of others*. Ultimately, the aim of this chapter is to provide readers with some empirical food for thoughts regarding what can be viewed as a major developmental transition from the evidence of self-awareness manifested at birth, and the awareness of the self as perceived by others, which starts to be manifested by the end of the second year in the life of the healthy developing child.

1. Public and private re-presentation of the self: *The Irreconcilable*

The 19th century French poet Arthur Rimbaud wrote as a young adult: “I is someone else” (“Je est un autre”). Rimbaud points to the difficulty of reconciling self-knowledge with the self that is known by others, what can be construed as the *irreconcilable difference between what we represent of ourselves privately and what is represented by others based on what we display publicly*. There is indeed, at least for adults, an inseparability and also a marked gap between what we perceive of ourselves and what we construe as the perception of our-

selves by others. Typically, and this seems to be a universal psychological issue across cultures and for individuals of all walks of life, we can never quite reconcile what we feel about ourselves from within our own body, with what we construe as the perception of ourselves by others from their own embodied perspective.

As a case in point, let me mention the observations reported some 30 years ago by the visual anthropologist Edmund Carpenter. Carpenter recorded the reactions of adult Biami, an isolated tribe living in the plateau of Papua New Guinea, when introduced for the first time to their own mirror reflection, video image, and other Polaroid photographs of themselves. Carpenter reported a powerful expression of fear and anxiety in these adult individuals: “They were paralyzed: after their first startled response – covering their mouths and ducking their heads – they stood transfixed, staring at their images, only their stomach muscles betraying great tension” (Carpenter 1975: 452).

It is my contention that the fear expressed by the Biami adults, confronted for presumably the first time with their specular image, rests on the fact that the individual comes to grip with the profound discrepancy between what he or she feels and perceives about the self from within his/her own body, and what others actually perceive of him or her as it is objectified by the specular, photographic, or video image. In short, what the adult comes to grip with is the basic discrepancy between private and public self, a discrepancy that, I propose, is never filled. In development, the awareness of this discrepancy emerges by the end of the second year, with young children beginning to manifest embarrassment in front of mirrors, in addition to self-recognition.

2. Specular image and levels of self-awareness

Developmental psychologists have used the reaction of children to their own specular image as an index of self concept (Amsterdam 1972; Lewis & Brooks-Gunn 1979; Povinelli 2001), for example, when toddlers begin to reach for a dab of rouge surreptitiously applied on their face, which they have discovered in the mirror. This response, in the context of the so-called “rouge test”, is taken as indexing mirror self-recognition (Bertenthal & Fisher 1978). In fact, upstream and downstream of mirror self-recognition, as indexed by the passing of the rouge test, one can distinguish at least six basic possibilities of what an individual can perceive when seeing him- or herself in a mirror. I construe below that these six possibilities map onto five levels of self-awareness (six levels if we include a Level 0 of non self-awareness) as they unfold in early development

(see also Rochat 2003). This ontogenetic unfolding includes the mirror self-recognition, which is indexed by the passing of the rouge test, but also points to implicit levels of self-awareness that precede the passing of such a test, as well as explicit levels of self-awareness that follow it. I review these possibilities below in the order of their emergence, asking the question: what is the range of possible perceptions of the self in the mirror?

Possibility 1: The first possibility is that the individual perceives the specular image as *undifferentiated from other entities perceived in the environment*. The image of the self reflected in the mirror is confounded with the environment itself, hence the reflective property of the mirror is overlooked by the individual. The mirror is perceived as a mere extension of the surrounding space. Birds flying in a room with a mirror and crashing repeatedly against it express such undifferentiated perception. The individual perceived in the mirror stands for somebody else. Other birds, cats, dogs, or macaque monkeys who display aggressive behaviors to their specular image also manifest such possibility, somehow oblivious of the self-reflecting nature of the mirror (Zazzo 1981).

Possibility 2: A second possibility is that the individual perceives the specular image as *differentiated from other entities in the environment*. Instead of a mere extension of the surrounding space, the mirror reflection is detected as a solid, flat, polished surface that is differentiated from the three-dimensional layout of surfaces and objects. The mirror is seen as an object among other objects in the environment.

Possibility 3: A third possibility is that the individual perceives the specular image as differentiated from other entities, but also as *indexing invariant contingent relations between self-produced and seen movements*. The individual picks up the specificity of the specular image as temporally contingent and spatially congruent correspondences between proprioception and vision of his/her own body. This detection is perceptual, based on the detection of intermodal invariants in the mirror that specify the body as to how it looks and feels from within when set in motion. This third possibility is expressed, for example, by chimpanzees caught picking their teeth while guiding their action visually via a mirror (Povinelli 1993, 1995).

Possibility 4: The fourth possibility is that individuals actually recognize themselves in the specular image, as indexed by self-referencing in the context of the rouge test. The act of bringing the hand in contact with the spot of rouge on the face indicates that the child goes beyond the mere perception of the specular image as a perceptually differentiated entity in the world. The child manifests an explicit understanding that *what is reflected in the mirror*

stands for his/her own body that is facing it. The mirror is perceived as self-referencing; it refers and maps onto the currently felt body. The specular image is thus recognized as standing for the self, beyond mere intermodal contingency detection.

Possibility 5: The fifth possibility is that the individuals recognize themselves in the specular image whether it is contingent or not. Such recognition does not depend on the perfect contingency and spatial congruence between proprioception of the body and the vision of it moving as perceived in the present. The perceived image of *the body is identified as a permanent entity beyond the here and now of intermodal perception*, recognized as an image that refers to the self in the past, as for example in a photograph or in a pre-recorded video (Lewis & Brooks-Gunn 1979; Zazzo 1981; Povinelli 2001).

Possibility 6: The sixth possibility is that individuals not only recognize themselves in the specular image as permanent entities, but also see this image as standing for *what other people actually perceive of the self*. The specular image is recognized as standing for the private as well as the *public* self. This recognition opens the process of an exploration and ultimately an evaluation of the gap between how individuals feel from within as well as represent themselves privately, and how they are actually perceived by others as indexed by the mirror reflection. The specular image is construed as an objectified, publicly accessible self presentation that is open for evaluation and re-assessment.

3. Early development of self-awareness

The order of the 6 possibilities outlined above corresponds to an increasing complexity in the perception and representation of mirror self-reflection: from mere perceptual discrimination to explicit recognition, identification, and ultimately to evaluation. This progressive order corresponds also, I will contend, to the order in which self-awareness develops between birth and 3 years of age. Next, I present this development in its chronology and as it appears to unfold in the first 3 years of life. For illustration, I selected empirical evidence that support each of these proposed developmental steps. Table 1 below summarizes these steps, in the chronology of their manifestation early in life, with approximate age onset. Note that these ages are only indicative, varying across individual infants and depending on the experimental and cultural context of the child, as well as the kind of tests used to probe the development of self-awareness. Rather than exact timing, what is important here is the putative

Table 1. Proposed developmental progression in levels of self-awareness

Development Age		Process	Behavioral expression
Step 1	(possibility 1) —	confusion	(self-world fusion)
Step 2	(possibility 2) Birth	differentiation	(self-world discrimination)
Step 3	(possibility 3) 2–7 months	causation	(self-exploration)
Step 4	(possibility 4) 18 months	recognition	(self-objectification)
Step 5	(possibility 5) 24 months	extension	(permanence)
Step 6	(possibility 6) 36 months	evaluation	(co-awareness)

invariant developmental progression in terms of emerging processes, behavioral expression, and motives that seem to underlie the ontogeny of self-awareness.

3.1 Self-world differentiation at birth

Recent research shows that from the first minutes of life outside the womb, babies manifest a sense of their own bodies as differentiated entities among other entities in the environment (see Rochat 1997 for a more detailed discussion of experimental facts supporting this assertion). According to Neisser (1991, 1995; see also Rochat 1997), newborns manifest rudiments of a perceived or “ecological” self.

For example, in one study (Rochat & Hespos 1997) we showed that from birth, infants manifest a discrimination between tactile stimulation that is self produced (self-stimulation) and tactile stimulation from a non self, external origin (allo-stimulation). Comparing the rooting responses of newborns following a stimulation to either the right or left cheek, caused by either the finger of an experimenter (allo-stimulation) or the spontaneous transport of the infant’s own hand toward the face (self-stimulation), we observed that newborns tend to turn their head significantly more toward the experimenter’s finger compared to their own hand. When the hand of the experimenter was involved, the newborns showed more head orientation with mouth opening and sucking movements with tongue protrusion.

It appears that infants from birth are capable of discriminating information that specifies their own bodies as differentiated entities. This observation

is not trivial since it is contrary to the long held idea of an initial state of un-differentiation or confusion between infants and their environment (e.g., Piaget 1936). Some psychoanalysts went as far as elaborating theories of personality development on the premise that the starting point of such development is an initial state of un-differentiation or “infantile autism” (Mahler, Pine, & Bergman 1975).

Recent research indicates that, on the contrary, early on, infants process intermodal (polysensory) information that specifies the body as a distinct entity. Researchers have now accumulated numerous data demonstrating the remarkable coordination at birth of visual and postural/vestibular systems. Such coordination allows an infant to pick up information that specifies movements of his/her own body in a stable environment or the reverse, the stability of the body in a moving environment (Butterworth 1995; Bertenthal & Rose 1995; Jouen & Gapenne 1995). Such discrimination, which is based on the processing of perceptual information from multiple modalities, is evident from birth and probably the result of an active prenatal calibration of sensory and motor systems. Fine ultrasonic scanning of fetuses during the last 3 months of pregnancy reveals indeed that most of the behaviors observed immediately after birth in newborns are already functional and well established in the womb (Hopkins & Prechtl 1984; De Vries, Visser, & Prechtl 1984).

There is a remarkable continuity between pre- and post-natal behaviors (Prechtl 1984). This continuity suggests that the implicit knowledge of the body as differentiated entity expressed in newborns' behavior could well be the product of prenatal learning, as in the case of maternal voice discrimination expressed by newborns immediately after birth (DeCasper & Fifer 1980; DeCasper, Lecanuet, Busnell et al. 1994) or the evidence of neonates' olfactory discrimination of maternal amniotic fluid compared to the amniotic fluid of a female stranger (Marlier, Schaal, & Soussignan 1998).

The perceptual learning of their own bodies as differentiated from other entities in the world is the main pillar of an ecological sense of self expressed by infants from birth. Although far from a conception of the self as perceived by others, this basic sense of self is a necessary precursor, a *sine qua non* condition for the emergence of co-awareness. Questions remain, however, as to how infants develop co-awareness from this basic, early (perceived) sense of self.

3.2 Emerging intersubjectivity and self-exploration at 2 months

One can observe a radical behavioral reorganization with the apparition of the social smile at around 6 weeks of age. This reorganization corresponds to a

revolution in the way infants relate to the world, in particular, how they relate to others via reciprocal exchanges. This revolution is *de facto* the true psychological birth of the infant, the beginning of a sense of shared experience with others, hence the beginning of co-awareness (Rochat 2001).

For parents, witnessing the first smile of their child in the context of intimate face-to-face exchanges (as opposed to the automatic smile expressed by neonates during sleep or following feeding) is a major event. Nothing can exaggerate the importance of the emergence of socially elicited smiling in the life of a child and his caretakers. This emergence marks the beginning of the child's relational existence, as it is the first explicit manifestation of a shared positive experience. It is the first unmistakable manifestation of an experience of well being with others. It is also the first message of reciprocity that is not solely linked to basic physical care dispensed by the adult. It is a first message that begins a lifelong conversation with others. With socially-oriented smiling, infants affirm their presence in the world *with* others. It is the beginning of co-awareness and indeed the true psychological birth of the child.

Parallel to the emergence of social smiling, many other aspects of infants' behavior are reorganized. For example, during the second month, the capacity of infants' attention changes markedly and in a relatively sudden fashion. Wolff (1987) observed that by 6 weeks, infants spend significantly more time in an awake and alert state, spending significantly more time attending to their environment with eyes wide open. It is also by this age that infants begin to scan faces by focusing markedly more on the eyes and the mouth, facial regions that are rich with information regarding the fluctuating emotional states of others. In sharp contrast, neonates tend to focus much more on the periphery of the head (Maurer & Salapatek 1976; Haith, Bergman, & Moore 1977; see also the relevant work of Morton & Johnson 1991).

At the level of general cognitive development, the second month marks a change in the stance the infants take toward the world that surrounds them. There is some kind of a radical world view change. From birth, and even prior, infants are capable of complex sensory-motor learning and perceptual discrimination. However, this learning and discrimination are not yet under anything that resembles voluntary control, still dependent on the here and now or immediacy of perceptual experiences. There is not yet clear evidence of systematic groping or exploration. For example, numerous research studies done in the past 30 years demonstrate the stunning capacity of neonates to imitate facial expressions such as mouth opening, tongue protrusion, and even emotional facial displays such as happy or sad expressions (Meltzoff & Moore 1977; Field et al. 1982). However, this imitation is still rather fragile. It is not very system-

atic and does not show much flexibility. This led some critics to view neonatal imitation as nothing more than the product of innate automatic release mechanisms (Anisfeld 1991). By 6 weeks, infants' imitative behavior eludes such interpretation, clearly demonstrating that there is more to it than a pre-wired automatism. Meltzoff and Moore (1992) showed that by this age, infants begin to systematically modify their imitative response to match the adult model. For example, if the experimenter pulls his tongue to the side, the infant might first pull her tongue to the center and slowly bring it to the side to match the target gesture. This behavior shows systematic approximation and what amounts to willful groping.

Recently, we made similar observations comparing newborns' and 2-month-olds' sucking behavior on "musical" rubber nipples. In this research (Rochat & Striano 1999), every pressure applied by the infant on the nipple was associated with a contingent succession of sounds that were more or less the auditory analog of the oral pressures generated by the infant on the pacifier. In one condition (analog), the pitch variation of the successive sounds heard by the infant was proportional to the variations of pressures applied by the infant on the pacifier. In another (non-analog) condition, the pitch variation of the sounds was not dependent, but rather varying randomly. We observed that by 2 months, infants manifest a differential modulation of their suction of the pacifier depending on the condition (i.e., analog or non-analog auditory consequence of sucking). In contrast, we tested newborns, who did not show any evidence of such differential responding, hence no evidence of systematic exploration of the auditory consequences of their own oral (sucking) activities.

Around 6 weeks of age, babies thus manifest a novel stance toward objects, toward themselves, and toward others. This novel stance is a *contemplative* and *reciprocal stance*, as opposed to the discriminatory and immediate stance of newborns (Rochat 2001). This new stance is linked to expectations and the systematic exploration of physical events, as well as to the first reciprocal exchanges with others. Affective reciprocity by the second month is a major step toward co-awareness, a *sine qua non* condition of his/her development. From birth, which was the first *sine qua non* condition for the emergence co-awareness.

3.3 First signs of self-objectification by 18 months

Until the middle of the second year, when linguistic and symbolic competencies start to play a major role in the psychic life of children, self-awareness remains *implicit*. It is expressed in perception and action, not yet expressed via sym-

bolic means such as words. Prior to approximately 14–18 months, there is yet no clear evidence that children perceive traces of themselves, as *standing for* themselves, only themselves, and no one else, such as the little footprints they might leave in the mud or the image they see in the mirror.

However, months earlier, infants do discriminate between their own image and the image of another infant. Preferential looking studies show that by 5–6 months infants tend to be significantly more captivated by a pre-recorded video of another, same-age infant, compared to a pre-recorded video of themselves wearing an identical, same color outfit (Bahrick, Moss, & Fadil 1996). It appears that by this age, and presumably via previous exposure to mirrors and other self reflecting devices, infants pick up invariant features of their own faces. It does not mean however that they construe these features as standing for themselves. It is the product of perceptual learning of subtle invariant facial features they quickly become familiar with. When placed in a situation where they have the choice to explore either their own familiar face or the face of another child, they show a typical preference for novelty (e.g., Fantz 1964; Rochat 2001). Although certainly a necessary precursor and a sign of remarkable perceptual learning ability, this preference does not mean yet that infants do *recognize* that it is themselves on the TV.

The same kind of interpretation applies to our finding that 4- and 7-month-old babies show clear discrimination between seeing themselves live on a TV while moving around in their seat versus seeing a live experimenter on a TV engaged in the systematic imitation of what the infant is doing (Rochat & Striano 2002). In this experiment, the experimenter shadowed the infant as mirrors do. We found that infants smiled, vocalized, and looked differentially at the imitating experimenter seen on TV compared to the self. In addition, infants tended to react differentially depending on the condition when the image was suddenly frozen in “still-face” episodes.

In all, young infants demonstrated once again their perceptual ability to distinguish between the familiar sight of themselves and the novelty of the experimenter appearing on the TV, the age variable, notwithstanding the inescapable lack of perfect contingency in the Experimenter’s shadowing of the infant’s own actions (see Rochat & Striano 2001, 2002).

Despite all this perceptual discriminability between what pertains to the self and what pertains to others, up to the middle of the first year, infants appear oblivious that some rouge has surreptitiously been smeared on their faces or that a yellow “Post-It” might appear on their foreheads when looking at their own specular image (Bertenthal & Fisher 1978; Povinelli 1995). It is only by 18 months that infants start to reach for the mark on their own bodies, of-

ten in order to remove it. This behavior is considered by most developmental and comparative psychologists as the Litmus test of self-awareness. It is often viewed as evidence of a conceptual or “represented” sense of self in any organism that behaves like this in front of mirrors, whether human children, non-human primates, avians, mammals like elephants, or even cetaceans like dolphins (Parker, Mitchell, & Boccia 1994). But why is this? It is mainly because by showing this behavior, individuals demonstrate the ability to refer to the specular image as standing for their own bodies. In other words, they reference the silhouette they see reflected in the mirror to precise regions of their own bodies they cannot see directly (e.g., their foreheads). This would be impossible without a body schema or own body representation that is mapped onto what is seen in the mirror. Therefore, this behavior indicates that the mirror reflection is seen by the individual as standing for this representation. It is identified as referring to the body experienced and represented from within, not anybody else’s. Identity is used here in the literal, dictionary sense of “recognizing the condition of being oneself, not another” (Random House Unabridged dictionary, Second Edition, 1993).

In relation to the above formulation, mirror self-recognition expressed via the “successful” passing of the mark test is predictably linked to major progress in symbolic (referential) functioning of the child in other domains, in particular language development.

By 18 months, infants also start to mark contrasts between themselves and other people in their verbal production. They express semantic roles that can be taken either by themselves or by others (Bates 1990). An explicit, hence reflective conception of the self is apparent at the early stage of language acquisition, at around the same age that infants begin to recognize themselves in mirrors. This chronological link in development provides indirect validation of the mirror test and the interpretation I provided above. Indeed, as argued by Bates (1990), language acquisition requires a preexisting conceptual or represented sense of self as “Me” as opposed to simply “I”: “a theory of the self as distinct from other people, and a theory of the self from the point of view of one’s conversational partners” (Bates 1990: 165).

3.4 Developing self permanence by 24 months

If infants identify themselves in mirrors starting at 18 months, they still demonstrate that the *Me* they identify in the specular image remains enigmatic and ambivalent. They appear to still oscillate between an awareness of the self and an awareness of seeing someone else facing them (Piaget 1962;

Povinelli 2001; Rochat 2001). Identifying oneself in the mirror is a major feat, not only for the referential mapping between the mirror reflection and the own body schema, but also because what the child sees in the mirror is the way he or she always sees others: in an “en face” posture often with eye contact. In relation to this basic experience of social encounters, what the child experiences in the mirror might be “Me”, but it is also what others typically look like. The child therefore has to suspend and override his overall visual experience of others, the specular image standing for “Me as an other” (Me but Not Me dilemma, Rochat 2001; see also from a psychoanalytical perspective Jacques Lacan’s account of “the mirror stage”).

The mirror experience of the self carries this fundamental ambiguity and children struggle with it, as we will see, until at least their fourth birthday. Note that this ambiguity is pervasive all through the life span. As adults, we look at ourselves in mirrors, working on our presentation by simulating or representing the looking of others at our own bodies. What we are seeing, is *de facto* our appearance as seen by others, hence the pretense of someone else.

In his seminal observations of his own children, Piaget reported anecdotes that pertain to the mirror dilemma. Jacqueline, aged 23 months, announces to her father as they are coming back from a walk that she is going to see her father, her aunt, and herself in the mirror. Perfectly capable of identifying herself in the mirror as “Me” when prompted by her father asking “who is there?”, Piaget observes that Jacqueline provides also, at times, a *third person* account of what she sees in the specular image. Likewise, she tends to oscillate between claiming that it is “Me” or that it is “Jacqueline” when viewing photographs with herself in it (Piaget 1962:224–225).

More recently, as part of a series of ingenious studies on the developmental origins of self-recognition, Povinelli reported the commentary of a 3 year-old with a sticker on her forehead, viewing herself on a TV. She says: “it’s Jennifer...it’s a sticker” and then adds, “but why is she wearing my shirt?” (Povinelli 2001:81).

In all, these observations illustrate once again the Me-But-Not-Me dilemma, which children struggle with months after they show signs of self-identification in mirrors. Povinelli’s research demonstrates that children continue to struggle with well into their third year with the Me-But-Not-Me dilemma when viewing live or pre-recorded videos of themselves. For example, children 3 years and younger do tend to reach for a large sticker they see on top of their own heads while viewing a live video of themselves. In contrast, they don’t when viewing the replay of the same video taken only 3 minutes prior. Furthermore, when asked who was on the TV, it is only by 4 years that

the majority of children say “Me” rather than their proper name suggesting a first person stance rather than a third (see Povinelli 1995, 2001, for a review and discussion of this research).

The careful empirical work of Povinelli and colleagues on delayed self-recognition shows that it is not prior to approximately 3 years that children begin to grasp the temporal dimension of the self. They start to grasp that the self pertains not only to what is experienced now, but also to what was experienced then, what can be seen in a mirror now or in a movie tomorrow: the same enduring self.

3.5 Others in mind by 36 months and older

By the time young children begin to express and recognize themselves as enduring entities, they also begin to show major advances in their understanding of others. By 4–5 years, children demonstrate the ability of holding multiple representations and perspectives on objects and people. They can infer for example the particular age, relative sentience, temperament, and emotionality of a person by merely looking at the quality of a simple drawing. By this age, children infer the mind and affect of the artist behind a graphic symbol (Callaghan & Rochat 2003). This ability is linked to the developing child’s ability to construe false belief in others, as well as to grasp the representational status of graphic and other symbolic artifacts such as maps, photos, or scale models (Callaghan & Rochat 2003; Rochat & Callaghan, *in press*; DeLoache 1991; Olson & Cambell 1993; Perner 1991).

The development of representational abilities in general and theories of mind in particular corresponds also to evidence of meta-awareness in relation to the self. For example, when children begin to understand explicitly that another person holds a false belief, they necessarily understand that they themselves hold the right belief. In the same way, when infants demonstrate some construal of object permanence, they also demonstrate their own permanence in relation to objects (Rochat 2001). These terms are inseparable.

The expression of embarrassment in front of mirrors by 2–3 years can be interpreted as the first sign of young children’s awareness of their public appearance and how others perceive them. As proposed earlier, by this age, children begin to experience the basic fear-generating realization of a gap between how they perceive themselves from within and what people actually perceive from the outside.

An alternative interpretation would be that young children shy away from their reflections in the mirror, not because they are “self-conscious”, but rather

because they wrongly construe the presence of another child staring at them with some kind of a persistent still face. But this is doubtful considering, as we have seen, that very early on infants discriminate between seeing themselves or seeing someone else in a video (Bahrick et al. 1996; Rochat & Striano 2002).

By showing embarrassment and other so-called secondary emotions (Lewis 1992), young children demonstrate a propensity toward an evaluation of the self in relation to the social world. They begin to have others in mind, existing “through” in addition to “with” others.

Secondary emotions such as the embarrassment children begin to express by 2–3 years parallel and are probably linked to the emergence of symbolic and pretend play. Such play entails, if not at the beginning, at least by 3–4 years, some ability to simulate events and roles, to take and elaborate on the perspective of others (Harris 1991; Tomasello 1999; Tomasello, Striano, & Rochat 1999; Striano, Tomasello, & Rochat 2001).

The process of imagining what others might perceive or judge about the self, whether this imagination is implicitly or explicitly expressed, is linked to the cognitive ability of running a simulation of others’ minds as they encounter the self. There is fantasy and phantasms involved, the stuff that feeds the self-conscious mind and characterizes the meta-cognitive level of self-awareness. Note that the articulation in development between the evaluative sense of self expressed at level 5 via embarrassment and the meta cognitive awareness of level 6 remains for the most part a mystery and deserves much more experimental scrutiny.

4. The development of co-awareness: Toward a collaborative and seductive stance

The reference to others’ views starts a process that rapidly becomes a major determinant of infants’ and toddlers’ behavior. It leads the child toward a growing awareness of the self in relation to others. At the level of behavior, the emergence of this novel (self-conscious) awareness manifests itself most blatantly in the form of a proactive and systematic enterprise of seduction (in the general sense of enticing the attention of others for pleasure and comfort) leading the child to behave in increasingly irrational and phantasmal ways. It is the dawn of the complex nod of representations that children generate as to how they relate to others, how they are perceived, and ultimately valued by them. These representations range from the longed for sense of being loved and affiliated to the most dreaded sense of being rejected and disenfranchised.

The elaboration of these representations brings the social dependence of the young child to new, much more complex levels of meaning. These new levels of meaning are indexed by the blossoming of behaviors that defy reason and common sense. These include coy behavior, embarrassment, excessive and defiant behavior, irrational fears and anxieties in pretend play as well as in the form of nightmares during sleep.

At the level of exchanges with others, this psychological “revolution” also translates into the emergence of a whole range of proactive behaviors driven by the irresistible need to maintain affective proximity with others. This marks the beginning of young children’s active and selective attempts at engaging the people they encounter, rather than the reverse (adults actively, and often selectively engaging infants), which up to this stage has dominated their lives. As we know, games of seduction often defy reason! It is in this sense that parallel to the progress in logic and the rational conception of the physical world that continues to be documented by numerous studies in the cognitive developmental tradition of Jean Piaget, by the second year, children develop also, and probably more decisively, a capacity for seduction that leads them to irrationality.

This development pertains to a world that is essentially subjective and phantasmal. It is the represented world regarding how others perceive, value, and eventually judge us.

Beyond their first birthdays, infants manifest a dependence toward others that defies more and more common sense and straightforward understanding. When their child begins to walk and even to run, it is common for parents to notice how toddlers seem systematically attracted by the most dangerous obstacles in the environment: stairs, roads, furnaces, and other threatening features. These kinds of behaviors become quickly a means by which infants express defiance and gain renewed attention from the caretakers by controlling their panic intervention. Under the threat of defiant behaviors, parents are coerced into the undivided attention and exclusivity the infant is longing for.

Undivided attention of others on the self is indeed the ultimate expression of closeness and affective fusion that the young child is now actively looking for in others. Defiant behaviors mark the beginning of active seduction as a process of appropriation of others, in particular the appropriation of their undivided love and attention. In this process, children begin actively and systematically to coerce others into co-awareness. Note that this process is not unlike caretakers’ drive to coerce younger infants’ attention and positive emotions in silly games in an attempt to create a sense of shared experience.

To illustrate and give some empirical background for this developmental account, I report below three observations that point to the beginning of active

seduction at around the first birthday. In a study on the developmental origins of instructional learning, we recently examined the impact of the presence and interventions of others in a problem solving situation with various levels of difficulty (Goubet, Leblond, Poss, & Rochat 2001; Rochat et al. 2002). We systematically observed infants, aged between 9 and 18 months, presented with an attractive toy placed at a distance on a blanket in front of them. The infant sat on her mother's lap and an experimenter sat to the right of the infant. To grasp the toy, the infant first had to pull the blanket toward her to bring it within reach, a classic Piagetian means-end task that is solved at around 8 months (Piaget 1936; Frye 1995).

Our observations confirmed that the great majority of 9 month-old infants managed with no hesitation to pull the blanket and bring the toy toward them for further exploration and play. Curiously and rather unexpectedly, we found that this simple means-end performance tends to *deteriorate* by 14 and 18 months! At these older ages, about half of the infants do not try to pull the blanket. Rather, they desperately try to reach directly toward the distal toy by stretching and whining while looking at the experimenter. They request help and do not even seem to consider that they could manage to get to the object on their own. This behavior defies reason and does not reflect what infants at this age and following Piaget's account are clearly capable of doing in terms of means-end coordination. In fact, it appears that the physical meaning of a simple means-end task is now transformed into a more complex social and relational problem. It is as if others rather than the toy are becoming the game's end. The infant seems to construe the task as an opportunity to gain proximity and the undivided attention from others. The goal of the child is to commune and ascertain closeness with others, not to get to the toy. By the middle of the second year, the toy becomes a means to a social end, the end of creating co-awareness.

Another example indexing the emergence of an active process of seduction by the second year is illustrated with another observation we made of infants aged 9, 11, 14, and 18 months. Infants were facing an experimenter who systematically imitated the kind of actions they spontaneously performed on a toy (Agnetta & Rochat, in press; see also the original study reported by Meltzoff 1990). By 11 months, but particularly by 18 months, infants begin systematically to test the imitation of the experimenter by accelerating or suddenly stopping their own actions while staring at the experimenter and sometime smiling toward her. With this subtle mutual imitation game, infants attempt to ascertain their control of the experimenter's behavior by probing imitative responses. Again, with these actions, infants convey a sense of co-awareness.

They play on the same key with the experimenter, equally engaged in trying to be the imitator rather than the imitated. With this kind of development, infants reach new, more reciprocal levels of affective fusion and complicity with others.

Finally, another clear piece of evidence of a major step toward co-awareness is the emergence of embarrassment at around 18 months of age. Already from 2–3 months, infants demonstrate behaviors that look like embarrassment (i.e., smile accompanied by gaze aversion), when, for example, encountering an unfamiliar person (Reddy 2000). However, it is by 14 months that infants begin to manifest social embarrassment in a predictable and marked way, not only in the context of protracted attention on the self by others, but also in the context of a task or performance that can be evaluated by others.

By 18 months, the young child begins to manifest explicitly that he can recognize himself in a mirror, trying for example to wipe a spot of rouge that has been surreptitiously put on his face and that he discovers in the mirror (Gallup 1982; Zazzo 1981; Lewis & Brooks-Gunn 1979). Interestingly, we have seen that, aside from explicit self-recognition as in the rouge task, by 2–3 years, children also manifest embarrassment in front of their own specular image. This behavioral manifestation is very complex and even paradoxical, from the hiding of the face with arms and hands, gaze aversion, or sudden acting out in an apparent attempt to distract from what is revealed in the mirror (Fontaine 1992). The emergence of these behaviors is linked to the development of co-awareness, in particular the awareness of others' view of the self. With embarrassment, children indicate that what they perceive in the mirror is not only an image that refers to themselves (the identified and conceptual "Me" according to William James), but also what others can see of the self (in other words, the "public and potentially evaluated Me").

The development of self-awareness opens the door to the development of self presentation based on the very complex and often highly irrational process of representing how others perceive and evaluate us. This process certainly contributes to the development each individual constructs, according to his or her circumstances, of a sense of moral conduct (i.e., a sense of what behavior is socially more or less acceptable) and of a sense of affiliation (i.e., a sense of being more or less accepted by others). It is also on the basis of this process that children learn to collaborate with others and are able to engage in a didactic (i.e., explicitly instructional) relationship, either as teacher or student, all of which is resting on co-awareness. More importantly, it is on the basis of this process that children begin their career as compulsive seducers, exploring and exploiting for better or for worse the affective resources of their social environment, endlessly foraging for intimacy, proximity, and group affiliation.

Acknowledgments

The ideas presented in this chapter were also discussed in a recent article by the author for a special issue of the journal *Consciousness and Cognition* (2003) on self-perception and action, as well as in a chapter by the same author for a book entitled "*Theories of Infant Development*" (Bremner & Slater, Eds., Blackwell Publisher, 2003).

References

- Agnetta, B. & P. Rochat (2004). Imitative games in 9-, 14- and 18-month-old infants. *Infancy*, 6(1), 12–38.
- Amsterdam, B. (1972). Mirror self-image reactions before age two. *Developmental Psychobiology*, 5, 297–305.
- Anisfeld, M. (1991). Neonatal imitation. *Developmental Review*, 11, 60–97.
- Amiel-Tison, C. & A. Grenier (1980). *Evaluation neurologique du nouveau-né et du nourrisson*. Paris: Masson.
- Bahrnick, L., L. Moss, & C. Fadil (1996). Development of visual self-recognition in infancy. *Ecological Psychology*, 8, 189–208.
- Bates, E. (1990). Language about me and you: Pronominal reference and the emerging concept of self. In D. Cicchetti & M. Beeghly (Eds.), *The self in transition: infancy to childhood* (pp. 165–182). Chicago: University of Chicago Press.
- Bertenthal, B. & K. Fisher (1978). Development of self-recognition in the infant. *Developmental Psychology*, 14, 44–50.
- Bertenthal, B. I. & J. L. Rose (1995). Two modes of perceiving the self. In P. Rochat (Ed.), *The self in infancy: Theory and research* (pp. 303–326). Amsterdam: North-Holland, Elsevier Publishers.
- Butterworth, G. E. (1995). The self as an object of consciousness in infancy. In P. Rochat (Ed.), *The self in infancy: Theory and research* (pp. 35–52). Amsterdam: North-Holland, Elsevier Publishers.
- Callaghan, T. C. & P. Rochat (2003). Traces of the artist: Sensitivity to the role of the artist in children's pictorial reasoning. *British Journal of Developmental Psychology*, 21(3), 415–446.
- Carpenter, E. (1975). The tribal terror of self-awareness. In P. Hixson (Ed.), *Principles of visual anthropology* (pp. 56–78). The Hague, Netherlands: Mouton.
- DeCasper, A. J. & W. P. Fifer (1980). Of human bonding: Newborns prefer their mother's voices. *Science*, 208, 1174–1176.
- DeCasper, A. J., J. P. Lecanuet, M. C. Busnell, C. Granier Deferre, & R. Mangeais (1994). Fetal reactions to recurrent maternal speech. *Infant Behavior and Development*, 17(2), 159–164.
- DeLoache, J. S. (1991). Symbolic functioning in very young children: Understanding of pictures and models. *Child Development*, 62, 736–752.

- De Vries, J. I. P., G. H. A. Visser, & H. F. R. Prechtl (1984). Fetal motility in the first half of pregnancy. In H. F. R. Prechtl (Ed.), *Continuity of neural functions from prenatal to postnatal life* (pp. 46–64). Spastics International Medical Publications.
- Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, 146, 668–670.
- Field, T. M., R. Woodson, & R. Greenberg (1982). Discrimination and imitation of facial expressions by neonates. *Science*, 218, 179–181.
- Fontaine, A.-M. (1992). *L'enfant et son image*. Paris: Nathan.
- Frye, D. (1991). The origins of intention in infancy. In D. M. C. Frye (Ed.), *Children's theories of mind: Mental states and social understanding* (pp. 15–38). Hillsdale, NJ: Lawrence Erlbaum.
- Gallup, G. G. (1982). Self-awareness and the emergence of mind in primates. *American Journal of Primatology*, 2, 237–248.
- Gergely, G. & J. S. Watson (1999). Early social-emotional development: Contingency perception and the social-biofeedback model. In P. Rochat (Ed.), *Early social cognition* (pp. 101–136). Hillsdale, NJ: Erlbaum Publishers.
- Goffman, E. (1959). *The presentation of self in everyday life*. New York: Doubleday.
- Haith, M. M., T. Bergman, & M. J. Moore (1977). Eye contact and face scanning in early infancy. *Science*, 198, 853–855.
- Harris, P. (1991). The work of the imagination. In A. Whiten (Ed.), *Natural theories of mind* (pp. 283–304). Oxford: Blackwell.
- Hopkins, B. & H. F. R. Prechtl (1984). A qualitative approach to the development of movements during early infancy. In H. F. R. Prechtl (Ed.), *Continuity of neural functions from prenatal to postnatal life* (pp. 179–197). Oxford: Blackwell Scientific Publications.
- James, W. (1890). *Principles of psychology*. New York: Henry Holt.
- Jouen, F. & O. Gapenne (1995). Interactions between the vestibular and visual systems in the neonate. In P. Rochat (Ed.), *The self in infancy: Theory and research* (pp. 277–302). Amsterdam: North-Holland/Elsevier Science.
- Kagan, J. (1984). *The nature of the child*. New York: Basic Books.
- Lewis, M. (1992). *Shame: The exposed self*. New York: The Free Press.
- Lewis, M. & J. Brooks-Gunn (1979). *Social cognition and the acquisition self*. New York: Plenum Press.
- Loveland, K. A. (1986). Discovering the affordances of a reflecting surface. *Developmental Review*, 6, 1–24.
- Mahler, M. S., F. Pine, & A. Bergman (1975). *The psychological birth of the human infant*. New York: Basic Books.
- Marlier, L., B. Schaal, & R. Soussignan (1998). Neonatal responsiveness to the odor of amniotic and lacteal fluids: A test of perinatal chemosensory continuity. *Child Development*, 64, 611–623.
- Maurer, D. & P. Salapatek (1976). Developmental changes in the scanning of faces by young infants. *Child Development*, 47, 523–527.
- Meltzoff, A. N. & M. K. Moore (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198, 75–78.

- Meltzoff, A. N. (1990). Foundations for developing a concept of self: The role of imitation in relating self to other and the value of social mirroring, social modeling, and self practice in infancy. In D. B. M. Cicchetti (Ed.), *The self in transition: Infancy to childhood* (pp. 139–164). The John D. and Catherine T. MacArthur foundation series on mental health and development. Chicago, IL: University of Chicago Press.
- Meltzoff, A. N. & M. K. Moore (1995). A theory of the role of imitation in the emergence of self. In P. Rochat (Ed.), *The self in infancy: Theory and research* (pp. 73–94). Amsterdam: North-Holland, Elsevier Publishers.
- Meltzoff, A. N. & M. K. Moore (1992). Early imitation within a functional framework: The importance of person identity, movement, and development. *Infant Behavior & Development*, 15, 479–505.
- Morton, J. & M. H. Johnson (1991). CONSPEC and CONLERN: A two-process theory of infant face recognition. *Psychological Review*, 98, 161–181.
- Neisser, U. (1991). Two perceptually given aspects of the self and their development. *Developmental Review*, 11, 197–209.
- Neisser, U. (1995). Criteria for an ecological self. In P. Rochat (Ed.), *The self in infancy: Theory and research. Advances in psychology*, 112 (pp. 17–34). Amsterdam, Netherlands: North-Holland/Elsevier Science Publishers.
- Olson, D. & R. Cambell (1993). Constructing representations. In C. Pratt & A. F. Garton (Eds.), *Systems of representation in children: Development and use* (pp. 11–26). New York: Wiley & Sons.
- Parker, S. T., R. W. Mitchell, & M. L. Boccia (1994). *Self-awareness in animals and humans*. Cambridge: Cambridge University Press.
- Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.
- Piaget, J. (1936). *La naissance de l'intelligence*. Neuchâtel, Paris: Delachaux & Niestlé.
- Piaget, J. (1962). *Play, dreams and imitation in childhood*. New York: Norton.
- Povinelli, D. J. (1993). Reconstructing the evolution of mind. *American Psychologist*, 48, 493–509.
- Povinelli, D. J. (1995). The unduplicated self. In P. Rochat (Ed.), *The Self in Infancy: Theory and Research* (pp. 161–192). Amsterdam: North-Holland/Elsevier Science.
- Povinelli, D. J. (2001). The self: Elevated in consciousness and extended in time. In C. Moore & K. Lemmon (Eds.), *The self in time: Developmental perspectives* (pp. 75–95). Mahaw, NJ: Lawrence Erlbaum Associates.
- Prechtl, H. F. R. (Ed.). (1984). *Continuity of neural functions from prenatal to postnatal life*. Oxford: Blackwell Scientific Publications.
- Random House Unabridged dictionary, Second Edition (1993). New York: Random House.
- Reddy, V. (2000). Coyness in early infancy. *Developmental Science*, 3, 186–192.
- Rochat, P. (1995). Early objectification of the self. In P. Rochat (Ed.), *The self in infancy: Theory and Research* (pp. 53–71). Amsterdam: North Holland/Elsevier Science.
- Rochat, P. (1997). Early development of the ecological self. In C. Dent-Read & P. Zukow-Goldring (Eds.), *Evolving explanations of development* (pp. 91–122). Washington, DC: American Psychological Association.
- Rochat, P. (1992). Self-sitting and reaching in 5–8 month old infants: The impact of posture and its development on early eye-hand coordination. *Journal of Motor Behavior*, 24, 210–220.

- Rochat, P. (2001). *The infant's world*. Cambridge: Harvard University Press.
- Rochat, P. & N. Goubet (1995). Development of sitting and reaching in 5–6-month-old infants. *Infant Behavior and Development*, 18, 53–68.
- Rochat, P., N. Goubet, & S. J. Senders (1999). To reach or not to reach? Perception of body effectivities by young infants. *Infant and Child Development*, 8, 129–148.
- Rochat, P. & S. J. Hespos (1997). Differential rooting response by neonates: Evidence for an early sense of self. *Early Development & Parenting*, 6(150), 1–8.
- Rochat, P. & T. Striano (1999). Emerging self-exploration by 2 month-old infants. *Developmental Science*, 2, 206–218.
- Rochat, P., J. Querido, & T. Striano (1999). Emerging sensitivity to the timing and structure of protoconversation in early infancy. *Developmental Psychology*, 35, 950–957.
- Rochat, P. & T. Striano (1999b). Social cognitive development in the first year. In P. Rochat (Ed.), *Early Social Cognition* (pp. 3–34). Mahwah, NJ: Lawrence Erlbaum Associates.
- Rochat, P. & T. Striano (2001). Perceived self in infancy. *Infant Behavior and Development*, 23, 513–530.
- Rochat, P. & T. Striano (2002). Who is in the mirror: Self-other discrimination in specular images by 4- and 9-month-old infants. *Child Development*, 73, 35–46.
- Rochat, P. (2001b). Social contingency detection and infant development. 65. 3. *Bulletin of the Menninger Clinic* (pp. 347–361).
- Rochat, P. (2003). Five levels of self-awareness as they unfold early in development. *Consciousness and Cognition*, 12, 717–731.
- Stern, D. (1985). *The Interpersonal World of the infant*. New York: Basic Books.
- Striano, T., M. Tomasello, & P. Rochat (2001). Object and social support for early symbolic play. *Developmental Science*, 4(4), 442–456.
- Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.
- Tomasello, M., T. Striano, & P. Rochat (1999). Do young children use objects as symbols? *British Journal of Developmental Psychology*, 17, 563–584.
- Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. M. Bullowa (Ed.), *Before speech: The beginning of interpersonal communication* (pp. 321–347). New York: Cambridge University Press.
- Wolff, P. (1987). *The development of behavioral states and the expression of emotions in early infancy*. Chicago: The University of Chicago Press.
- Zazzo, R. (1981). Miroir, images, espaces. In P. Mounoud & A. Vinter (Eds.), *La reconnaissance de son image chez l'enfant et l'animal*. Collection Textes de Base en Psychologie (pp. 77–110). Neuchâtel: Delachaux et Niestlé.

CHAPTER 2

Threesome intersubjectivity in infancy

A contribution to the development of self-awareness

Elisabeth Fivaz-Depeursinge,¹ Nicolas Favez, and
France Frascarolo

University of Lausanne / University of Fribourg

Developmentalists see the origins of self-awareness as closely linked with intersubjectivity. Self-awareness is primarily social and co-constructed, with others from the outset providing the infant with “deep mirrors” by means of a reflecting process of her person (Rochat 1995: 69; and this volume). They distinguish between a secondary, reflective, self-awareness, which occurs when the self is the object of one’s own cognition, and a primary form of self-awareness, when the self is the object of one’s own perception (Butterworth 1995).

The same distinction is made regarding the development of intersubjectivity. From a developmental perspective, a nine month-old infant who expresses pleasure or signals frustration or uncertainty, in reference to an object on which she co-oriens with her mother, is making a bid for intersubjective communication (Trevarthen & Hubley 1978). These so-called triadic interactions emerging towards the end of the first year, point to secondary intersubjectivity, because they involve referential communication and the capacity to share attention and intentions with others (Tomasello 1995). A growing minority of authors makes a case for a primary form of intersubjectivity at an earlier age (Trevarthen 1984; Rochat & Striano 1999; Stern 2004). Infants appear to have an inborn or early capacity to share others’ feelings and mind-states. Findings in favor of the notion of primary intersubjectivity include neonatal imitation, whereby infants imitate and use non-verbal gestures, presumably with the purpose of identifying people (Meltzoff & Moore 1995). Young infants also engage in “protoconversations” and demonstrate a capacity to vary the timing and intensity of communication in unison with their partner. Furthermore, when mothers

mirror their infants' affect, the infants, by corollary, show a sensitivity to maternal affective mirroring. These results are evidence of primary intersubjectivity, namely a "sense of shared experience" (Rochat & Striano 1999). Thus, primary intersubjectivity characterizes early interactions that have no topic other than the interaction itself, with the partners affectively resonating to one another. Secondary intersubjectivity refers to interactions in which partners are intentionally exchanging messages about a common topic (Bretherton 1992).

The goal of this paper is to report new empirical evidence supporting intersubjectivity both at the level of the father-mother-infant triad, and in the primary domain. It involves the infant's early capacity for triangulation. Young infants are able to engage not only with one person at a time, but also with two, in this case their fathers and mothers. For instance, in a trilogy context, a three-month-old infant who is sharing pleasure or interest with one parent will spontaneously turn to the other one to share her affect with him or her too. We will consider the implications of the infant's triangular capacity for self-awareness in the discussion section.

1. Triangular interactions

We chose the term "triangular" for labeling three-person interactions to distinguish them from the "triadic" interactions between two persons about a physical object. Triadic capacities correspond to the emerging ability of the infant to monitor others in relation to objects (Rochat & Striano 1999:24). Triadic interactions have been extensively studied in developmental research as markers of the advent of the intentional stance and of secondary intersubjectivity (Bakeman & Adamson 1984; Trevarthen & Hubley 1978). But triangular interactions are also important for social cognition. As members of a social species, we are to grow and live in multiperson relationships. We know that infants are embedded more frequently in multiperson contexts than in strict dyadic interactions from the moment of birth (Dunn 1991; Schaffer 1984). Thus, it would be adaptive for them to be able to manage more than two-person relationships from the outset and to establish threesome or collective intersubjectivity among the members of a family. Hence, we assume, as does D. Stern, that intersubjectivity constitutes a motivational system in itself, in particular in the same capacity as attachment, as defined by Bowlby (1980). Intersubjectivity pertains to psychological belonging and intimacy. It is related to, but separate from, the attachment system, which specifically relates to physical closeness and group bonding. Intersubjectivity confers survival advantage

by promoting group formation, enhancing group functioning, and assuring group cohesion by giving rise to morality (Stern 2004).

It is important to note that triangular interactions are more complex than dyadic ones; they are enacted not solely in one context, but in four different contexts: not only the single “3-together”, where all partners are actively engaged, but also the three different “2+1s”, where two partners are actively engaged and the third one is in a third-party position. In other words, handling triangular interactions comes down to negotiating the four contexts considered as equivalent components within a composite system (Fivaz-Depeursinge & Corboz-Warnery 1999).

We came across the yet undescribed triangular capacity as we were studying how parents played conjointly rather than separately with their baby, comparing clinical families, with severe parental psychopathology, to ordinary ones. Longitudinal observations documented differences between infants growing up with parents in a harmonious coparenting relationship, namely the parents supporting each other in their parental function, versus parents in conflict. The former infants engaged more easily in triangular interactions and received more adjusted and sensitive responses from them than did the latter (Fivaz-Depeursinge & Corboz-Warnery 1999; McHale & Cowan 1996). The development of the capacity to handle triangular interactions presumably bodes well for the socialization of the child, for her understanding of multiperson relationships and her theory of mind.

2. The Lausanne trilogue play paradigm

The Lausanne trilogue play (LTP for short) is a semi-standardized situation in which one observes families as they proceed through the four contexts mentioned above: (1) one parent, e.g. the mother, and baby playing, the father being present as a third party (2+1); (2) father and baby playing, with the mother present as third party (2+1); (3) both parents playing with the baby (3-together); (4) father and mother talking together and baby present as third party (2+1). Because the natural goal of trilogue play, like of dialogue play, is to playfully share positive affect, this task primarily targets intersubjective sharing of affect. Technically, the setting of the LTP allows for the video-recording of the three partners facing each other in a triangular formation, with a sufficient resolution to allow for microanalytic coding of body and facial movement (see Fivaz-Depeursinge & Corboz-Warnery 1999).

The results reported in this paper are based on the microanalysis of interactions at half-second intervals. The sample included 40 volunteer families with medically intact first-borns, high to average socio-economic status, who were assessed at five points in time (pre-natally at the 5th month of pregnancy, post-natally at three, four, nine and 18 months). LTP sessions last about 8–10 minutes with younger infants, and about 10–15 minutes with older ones.

3. Triangular process in the domain of secondary intersubjectivity

Let us enter the documentation of triangular process at the developmental turning point of nine months, when triadic interactions are known to emerge. Perhaps one of the most striking non-verbal microprocesses revealing intersubjective communication is social referencing, whereby an infant facing the unknown consults her mother for guidance. This has been the focus of empirical research, for example, the interactions between mother and infant in puzzling situations or regarding puzzling objects at the end of the first year (Klennert, Campos, Sorce, Emde, & Svejda 1983). We also observed a triangular form of social referencing at nine months in trilogue play. These infant triangular bids evolved in four steps, lasting about 3–5 seconds, which is within the bounds of working memory (Baddeley 1992). First, surprised by an ambiguous event, the infant suddenly interrupts her activity and attends to the event. It may be an external event, e.g., a noise, or the facilitator leaving the room after giving the instructions; or an internal event between the partners, e.g., the mother proposing a new game. Second, the baby looks up at the parent with whom she is playing, with a quizzical expression, as if asking, “what is happening there?” She may or may not get a satisfying answer to her implicit question. Third, she turns with the same expression to the other parent and once again, may or may not get a satisfying answer. Fourth, she selects a course of action. I. Bretherton has convincingly argued that social referencing can be understood as one aspect of the infant’s ability to “interface minds” through intentional communication. Thus, infants have a sense of a shared mental world, a primitive ability to take on the role of the other (Bretherton 1992).

The parents’ ways of validating the babies’ bids revealed themselves essentially in two respects. First, in terms of their adjustment to the infant’s “question”, parents may read the infant’s signals with precision and offer reassurance, mind reading, or affect attunement; in other words, respond at the secondary intersubjective level. For instance, in affect attunement, instead of imitating the infant’s behavior *per se*, the parent matches the behavior’s tempo-

ral or spatial form, translating it into another modality. Thus, the parents focus the infants' attention on the shared affect rather than on the shared behavior (Stern 1985). But they may also miss the point, and distort the infant's communication, attune selectively, underattune, or overattune. They may plainly reject or ignore the infant's bids. Thus, parents' interpretations of the infants' bids offer infants not only a basis, but also a bias for the construction of social reality (Bretherton 1992). Second, and most important in triangular interactions, is the consistency of a parent's response with that of the other parent. This is crucial for enabling the child to make out a clear meaning of the event and select a course of action. This being a split-second event, the parents' responses are intuitive and mostly unaware. Imagine the consequences of the father giving the infant a reassuring look versus a contemptuous one when mother has proposed a new game. Triangular bids of social referencing may occur during the "2+1s" when the infant is playing with one parent and spontaneously turning to the other one in the third party position. In the same context, it is also likely that at a high point, when the infant is experiencing pleasure with say, his mother, he will spontaneously turn to his father, presumably to share this pleasure with him, to then turn once again to his mother. Sharing pleasure with or signaling protest to both parents are understandably most frequent in the "3-together" context. Finally, infants also address triangular bids when they are the onlookers of their parents' dialogue, be it only because they are puzzled by the change from being invited to engage to being left in the third party position, or to try to attract their parents' attention by means of charm or protest.

The microanalysis of these triangular events in our longitudinal sample of infants nine months old, showed that they made, on average, more than one triangular bid per minute for sharing pleasure, interest, protest, or for social referencing. As noted above, individual differences were important; they were partially accounted for by the degree to which parents coordinated with each other and with their child in play. Infants' bids were more numerous and more often positive when the parents responded in an adjusted way with affect attunement, mind reading, mirroring, or empathy. They were less frequent and more often negative when the parents tended to misunderstand, ignore, silence them or meet them with hostility. As mentioned previously, this was the case when problematic patterns of coparenting were observed, such as hostile-competitiveness or skewed engagement between parents when in relation to the child (Fivaz-Depeursinge, Frascarolo, & Darwiche 2002).

The triangular process being already in place by nine months, it is obviously forming before that age. Our observations of three and four month olds highlight a primary form of threesome intersubjectivity.

4. Triangular process in the domain of primary intersubjectivity

Let us first illustrate an early triangular process by means of a quasi-natural experiment of triangular invisible imitation – i.e., an infant matching proprioceptively actions she sees performed by others – as observed in a single case during the LTP.²

Manon, a three month old, was an energetic little girl. Her parents gave her vigorous and rapid stimulation and perceived her reactions with great precision. Manon liked it. She sustained a high level of attention throughout the interaction, even when she had to accumulate signals at her parents' address that they should lower the stimulation level. The story evolved in four episodes, one in each of the different parts of the LTP.

In the first part, Manon played with her mother, while her father was in the third party position. After a while, they started taking turns at sticking out their tongues. Manon imitated her mother with concentration and enthusiasm twice in a few seconds, and her mother congratulated her.

We know now that the neonate imitates various acts, such as tongue and lip protrusion, mouth and fist opening and closing, looking up, as well as several affect expressions (smiling, surprise, sadness). The studies on these phenomena are numerous, carefully done, replicable, and the divergences are more a matter of interpretation than of fact. One of these divergences concerns the reflexive versus intentional nature of these imitations. Meltzoff has convincingly demonstrated against the reflexive interpretation, by placing a pacifier in the infant's mouth during the presentation of the model; disturbed by the pacifier, the infant nevertheless imitated tongue protrusion after it was removed – an act that is incompatible with reflexive and automatic behavior. It involves, on the contrary, an intentional mapping between self and other (see Gallagher & Meltzoff 1996, for a review).

The details of the work involved in Manon's imitation also point to that issue: she began to look at her mother attentively, then she averted her eyes a bit; she waved her arms, took a look at the second presentation of the model, stuck out her tongue moderately, looked again at her mother while waving her arms, in a considerable effort of concentration; her mother stuck out her tongue a third time, and finally Manon stuck out her tongue more boldly, with concentration; delighted, she did it again and once more five seconds later, for good measure.

These observations fit a model of imitation as an active process, presumably innate. According to Meltzoff and Moore, infants use proprioceptive monitoring to correct their unseen imitation movements in order to bring them

into line with the visual target. It is an active intermodal mapping that allows the infant to represent acts within a common supramodal framework (Meltzoff & Moore 1995). It differentiates by steps and is at no moment devoid of a psychic function. On the basis of a series of sophisticated studies during these last twenty years, Meltzoff and Moore suggest that imitation serves as identification and communication with others. In the same way a baby looks at her hand and moves it in order to explore its possibilities by means of vision and proprioception, or manipulates a physical object in order to know it, the infant imitates her partner to get to know her (Meltzoff & Moore 1995).

In the second part of the LTP, Manon plays with her father and the mother takes the third party position. Manon initiated the tongue game with her father and they both had fun. He took her hands or her feet to make her pedal and responded to her facial signals with exaggerated mimics, the way adults do when playing with babies (Papousek & Papousek 1987). As father and daughter reached an optimal level of animation, Manon initiated the tongue game. But, having acknowledged her invitation with a smile, the father refused to take up that game; instead, he made a mock expression of disgust and explained with a smile: "It is not dads that show you how to stick out your tongue; moms are the ones to teach you that!" Thereupon, Manon, wary, gave him a half, mitigated smile, paused to suck on her fist, reoriented her gaze towards her father and they resumed playing. For her, this frustrating and puzzling experience presumably stood in contrast to her experience of reciprocity in the many games she had been playing with him before. What's more, she likes her behavior to be matched, as has been shown experimentally. Infants recognize being matched and increase their behavior when matched (Meltzoff & Moore 1995:83).

Assuming Manon has differentially identified her parents, we might observe the consequences of this distinction during the next part of the LTP, when she plays with both parents. Sometimes, families develop a theme across the four contexts of the LTP.

However, it should be noted, before examining this 3-together play, that Manon had played in "2+1" contexts up to this point. That is, she had played with one parent at a time, the other one remaining passive, but visible in the periphery. In spite of that, she had manifested that she considered these "2+1" as triangular contexts by addressing them with triangular bids. For instance, at the beginning of the first part, and several times after that along the way, she turned to look at father, the third party parent, and addressed him with the affective signal that she had just addressed to her mother, with whom she was playing. It could be a signal of interest, pleasure, or frustration. These

rapid transitions during which she shifted her attention and affect between her parents manifested her capacity to interact in a triangular way.

From the onset of the “3-together”, Manon was in a position to recognize that she now had two equal partners, thanks to the synchrony and coordination between the parents’ solicitations. She made a number of triangular bids, alternating her look between her father then her mother, then back to her father, etc., all the time shifting signals of interest, pleasure, or frustration between them.

The tongue game reappeared on Manon’s initiative when the parents, engrossed in their game, began puffing out their cheeks in synchrony. As her turn came to “reply”, Manon began by looking at father very attentively, and addressed him with the same half, mitigated smile she had addressed him with when he had refused to comply with the tongue game. Then she turned to mother and stuck her tongue out resolutely at her, looking at her straight in the eyes. The mother replied immediately by sticking out her tongue and one sensed from her eyebrow frown that Manon was getting ready to reply again. But she was interrupted in her momentum by her mother’s move towards the father, as the latter was stating in an adult tone of voice: “This is too complicated!”, and they laughed together. Manon turned to him with a look of interest and her father engaged in a series of exaggerated grimaces that were variations of the theme: closing his mouth in order to prevent himself from sticking his tongue out. Having looked at the father, the mother, laughing, called out to Manon: “What is dad doing...” and Manon having given a second half-smile to her father, turned towards her mother and stuck her tongue out at her again. The mother exclaimed: “This is right!”.

In other words, the little girl was acting as if selectively playing at sticking out her tongue at her mother and refraining from it with her father, in a triangular context in which her acts were coordinated. The parents themselves showed that they had in mind a conjoint scenario in which mother played the tongue game and father played at refraining from it.

The 3-together part terminated beautifully with a threesome vocalization game initiated by Manon. The little girl was tired and the parents soon decided to proceed to the last part in which she would be the third party. They engaged in an animated conversation while Manon looked attentively back and forth, visibly interested by what was going on between them. At one point, she resonated to their laughing together. Thus, she was also able to adopt the third party position and get interested in what was happening between her parents. Note that the effort put in by her throughout the play was not devoid of moments of fatigue or distress. At these times, Manon and her parents shared

these feelings, with the parents empathetically responding to her with affect attunement.

We have documented the phenomenon of triangular processes at 3 and 4 months in our longitudinal sample. At three months, the analysis of triangular bids showed that infants made on average close to two bids a minute for triangular sharing with their parents during an LTP session. Inter-individual differences were large, but the number of bids was related to the degree of coordination of the family over trilogue play, both at three months and later at nine months (Fivaz-Depeursinge & Corboz-Warnery 1999; Lavanchy 2002).

In order to test for infant triangular capacity in a more controlled way, we used another version of the LTP at the age of four months, the LTP with still face. It included a context with a parent posing with a still face – i.e., looking at the infant without talking, touching, or moving – based on Tronick's paradigm (Tronick, Adamson, Wise, & Brazelton 1978). The still-face context was introduced in order to test whether the infant, in the face of stress, would turn to the third party parent, in a precursor of social referencing. Results from the 31/40 families with four month-old babies, in which parents followed the instructions well enough, confirmed those found in the regular LTP at three months. The infants also made about two triangular bids a minute. The affect signals that they shifted from one parent to the other during these bids were coherent and context sensitive. Moreover, as expected, the infants displayed the well-known reactions to a parent posing stillfaced, i.e., distracting themselves to regulate their affect, checking up on the parent and trying to charm him or her into reengaging. They also turned more often to the third party parent with a tense/interested, sometimes negative expression, as if to request help in meeting the situation or making out its meaning. There was also a tendency to make more triangular bids in the still-face than in the regular 2+1 context (Fivaz-Depeursinge, Favez, Frascarolo, Lavanchy, & de Noni, *in press*).

5. Discussion and conclusion

Thus, starting from a primary form, triangular interactions and representations differentiate in steps that parallel the development of dyadic interactions and representations. Indeed, not only do the very same phenomena take place towards the end of the first year, as described, but also they again reemerge in a more elaborate form during the second year, with the advent of moral emotions and symbolic thought. For instance, take the case of the 18-month-old toddler during the LTP, in which the parents engage in a dialogue with each

other after having played with the toddler, leaving him in the third party position. At first, toddlers are likely to look back and forth between their parents with a puzzled expression. It then often happens that they take up again the games they had successfully engaged in with the parents in the previous parts of the LTP; in so doing, on the one hand, they regulate their state and enjoy playing on their own; and on the other hand, they attempt to draw the parents' attention, as attested by the triangular charm bids that they emit, as they reach the high points they had experienced in previous play with their parents. But sooner or later, they will test their parents insistently by requesting their attention, alternating between charm, provocation, and protest. In other words, toddlers show a more elaborate form of understanding of their parents' intentions and states of mind and try to alter them by all the means they have acquired.

Triangular bids will take an even fuller form in three to four year olds, when children begin co-constructing auto-biographical narratives with their parents. Words, as acts of meaning as defined by Austin (1962), will by then be fully embedded in bodily communication.

Within each of these domains of triangular processes, children reexperience in a new way their position in the different triangular contexts and enrich their representations of their sense of self with two others (Stern 2004).

Thus, the above observations and findings show that triangular capacity is already observed in infancy. Early triangular processes correspond to the key aspects of primary intersubjectivity as defined by Rochat and Striano:

The sense of shared experience that emerges from reciprocity...It entails a basic differentiation between self and other...and the capacity to compare and project one's own private experience onto another, e.g. "the like me stance".

(1999:5)

What's more, triangular interactions create dynamics not present in dyadic ones. Not only is the complexity multiplied by the addition of a third person, but the social feedback provided by this third party and the increased number of interactive contexts in which an infant can share her affect make for a much richer context.

In fact, the triangular bids of three to four month olds do not yet have the status they will have by the end of the first year, when they will exhibit the early markers of the intentional stance, such as joint attention, social referencing, and symbolic gestures (Bruner 1978). They are still immediate, based on a much tighter coupling between action and perception, and more context dependant (Butterworth 1998). Likewise, referential bids made at the end of the

first year do not yet have the status of self-reflective triangular bids that they will have with the advent of symbolic thought and the emergence of moral emotions in toddlerhood.

According to Rochat (1995), the reciprocity that characterizes dyadic and triangular interactions is instrumental in scaffolding self-awareness. Although it is only during the second year that self-awareness will manifest itself unambiguously, when toddlers become explicit about their idea of “me”, self-awareness does develop in the course of the first months. If it is primarily social, then triangular capacities and the practice of triangular interactions must contribute to a large extent to the differentiation of self and others. This differentiation was highlighted in the context of Manon’s family imitation game, a quasi experiment that provided the opportunity to examine the self in a socially contextualized situation (Gallagher & Marcel 1999). As noted by Meltzoff and Moore (1995: 82), imitation is a naturally occurring behavior that provides a visible read-out of the infant’s notion of bodily self. One may assume that Manon experienced different selves, or variations of self, in the four different contexts of the LTP: having played sticking out the tongue with mother and made a “like me” experience, she then encountered her father’s resistance to engage in that particular game; this was a “not like me” experience, informing her of the relative efficacy of her own actions, and of her difference, yet in a context of playfulness and reciprocity. In the three-together context, the two experiences were then coordinated into an overarching game in which the partners appeared to share a joint representation. Finally, Manon experienced the third party position, watching her parents’ dialogue and resonating to it. Thus, she was able to actively probe her parents and got different responses in terms of mirroring and reciprocity. It should be emphasized that the entire experience occurred in a context of parental emotional scaffolding, playfulness, and empathy. Thus, the situation multiplied the possibilities for Manon’s understanding of the differentiation of self and others.

To be noted were the particular moments when she began to prepare for imitating, before performing. They evoke the “tryings”, described by Gallagher and Marcel (1999: 13), as providing the actor with an awareness of action and a sense of agency.

In sum, imitation and triangular affect sharing in a threesome entail a form of bodily self-awareness in relation to two others and a beginning understanding of “being a subject among subjects” (Nadel & Tremblay 1999: 204).

Primary triangular process and its corollary, threesome intersubjectivity, also constitute yet another argument for considering that infants are in possession of a form of bodily self-awareness long before they are in possession of a

theory of mind (Zahavi, this volume). In the domain of secondary intersubjectivity, infants then acquire an understanding that one mind can be interfaced with another, namely a rudimentary theory of mind or ability to impute mental states to self and other. However, this does not mean that they are able to reflect on their own theory of mind. They simply operate with it (Bretherton 1992). In this regard, the development of intersubjectivity, twosome and threesome, reveals a necessary element in the development of self-awareness and theory of mind. As noted by Rochat, affects are major determinants of behavior and are crucial for monitoring, prediction, and control of others' behaviors. The development of intersubjectivity allows the infant to know more accurately the behavior of those on whom she depends (Rochat 1999:9).

Thus, the triangular process provides from the outset a serious advantage for the development of self-awareness, offering a wide range of possibilities and providing for a wider early experience of an "embodied practice of mind" (Gallagher 2001).

Acknowledgment

This research was supported by the Swiss National Scientific Research Found grant No 32-52508.97.

The authors wish to express their gratitude to Gabriela Gesinus-Visser for editing the English version of this paper.

Notes

1. Correspondence concerning this article should be addressed to E. Fivaz-Depeursinge, Centre d'Etude de la Famille, Site de Cery, Ch 1008 Prilly-Lausanne, Switzerland. E-mail: Elisabeth.Fivaz@inst.hospvd.ch; phone: +41.21.643.6401; fax: +41.21.643.65.93.
2. This case was described in a previous paper, see Fivaz-Depeursinge (2001).

References

- Austin, J. L. (1975). *How to do things with words* (2nd ed.). Oxford: Clarendon Press.
- Baddeley, A. D. (1992). Working memory. *Science*, 255, 556–559.
- Bakeman, R. & L. B. Adamson (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, 55, 1278–1289.

- Bowlby, J. (1980). *Attachment and loss: Vol. 3. Loss: Sadness and depression*. New York: Basic Books.
- Bretherton, I. (1992). Social referencing, intentional communication, and the interfacing of minds in infancy. In S. Feinman (Ed.), *Social referencing and the construction of reality* (pp. 57–77). New York: Plenum Press.
- Bruner, J. S. (1978). From communication to language: A psychological perspective. In I. Marlova (Ed.), *The social context of language* (pp. 17–48). New York: John Wiley & Sons.
- Butterworth, G. (1998). Origins of joint visual attention in infancy [Commentary]. *Monographs of the Society for Research in Child Development*, 63, 144–166.
- Butterworth, G. (1995). The self as an object of consciousness in infancy. In P. Rochat (Ed.), *The self in infancy: Theory and research* (pp. 35–51). Amsterdam: Elsevier.
- Dunn, J. (1991). Young children's understanding of other people: Evidence from observations within the family. In D. Frye & C. Moore (Eds.), *Children's theories of mind* (pp. 97–114). Hillsdale, NJ: Erlbaum.
- Fivaz-Depeursinge, E. (2001). Corps et intersubjectivité. *Psychothérapies*, 21, 63–69.
- Fivaz-Depeursinge, E. & A. Corboz-Warnery (1999). *The primary triangle. A developmental systems view of mothers, fathers and infants*. New York: Basic Books.
- Fivaz-Depeursinge, E., N. Favez, C. Lavanchy, S. de Noni, & F. Frascarolo (in press). Four-month-olds make triangular bids to father and mother during Trilogue play with still-face.
- Fivaz-Depeursinge, E., F. Frascarolo, & J. Darwiche (2002, July). *The functions of infant triangular interactions in high versus low family alliances*. Paper presented at the World Association for Infant Mental Health conference, Amsterdam.
- Gallagher, S. & A. Meltzoff (1996). The earliest sense of self and others: Merleau-Ponty and recent developmental studies. *Philosophical Psychology*, 9, 213–236.
- Gallagher, S. (2001). The practice of mind. Theory, simulation, or primary interaction? *Journal of Consciousness Studies*, 5–7, 83–108.
- Gallagher, S. & A. Marcel (1999). The self in contextualized action. *Journal of Consciousness Studies*, 6, 4–30.
- Klinnert, M. D., J. J. Campos, J. F. Sorce, R. N. Emde, & M. Svejda (1983). Emotions as behavior regulators: Social referencing in infancy. In R. Plutchik & H. Kellerman (Eds.), *Emotion. Theory, research and experience* (pp. 57–86). New York: Academic Press.
- Lavanchy, C. (2002). *L'interaction visuelle de l'enfant de trois mois avec ses deux parents*. [Visual interaction of the three-month-old infant with both parents]. Unpublished manuscript, University of Geneva, Switzerland.
- McHale, J. & P. Cowan (1996). New directions in the study of family-level dynamics during infancy and early childhood. In J. McHale & P. Cowan (Eds.), *Understanding how family-level dynamics affect children's development: studies of two-parent families. New Directions for Child Development*, 74, 5–26.
- Meltzoff, A. N. & M. K. Moore (1995). A theory of the role of imitation in the emergence of self. In P. Rochat (Ed.), *The Self in Infancy: Theory and Research* (pp. 73–93). Amsterdam: Elsevier.

- Nadel, J. & H. Tremblay-Leveau (1999). Early perception of social contingencies and interpersonal intentionality: Dyadic and triadic paradigms. In P. Rochat (Ed.), *Early social cognition. Understanding others in the first months of life* (pp. 189–212). Mahwah, NJ: Erlbaum.
- Papousek, H. & M. Papousek (1987). Intuitive parenting: A dialectic counterpart to the infant's integrative competence. In J. D. Osofsky (Ed.), *Handbook of infant development* (2nd ed., pp. 669–720). New York: Wiley.
- Rochat, P. & T. Striano (1999). Social-cognitive development in the first year. In P. Rochat (Ed.), *Early social cognition. Understanding others in the first months of life* (pp. 3–34). Mahwah, NJ: Erlbaum.
- Rochat, P. (1995). Early objectification of the self. In P. Rochat (Ed.), *The self in infancy: Theory and research* (pp. 53–71). Amsterdam: Elsevier.
- Schaffer, H. R. (1984). *The child's entry into a social world*. London: Academic Press.
- Stern, D. N. (1985). *The interpersonal world of the infant*. New York: Basic Books.
- Stern, D. (2004). *The present moment in psychotherapy and everyday life*. New York: WW Norton.
- Tomasello, M. (1995). Understanding the Self as social agent. In P. Rochat (Ed.), *The self in infancy: Theory and research* (pp. 449–460). Amsterdam: Elsevier.
- Trevarthen, C. (1984). Emotions in infancy: Regulators of contact and relationships with persons. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 129–157). Hillsdale, NJ: Erlbaum.
- Trevarthen, C. & P. Hubley (1978). Secondary intersubjectivity: Confidence, confiding and acts of meaning in the first year. In A. Lock (Ed.), *Action, gesture and symbol. The emergence of language* (pp. 183–229). New York: Academic Press.
- Tronick, E., H. Als, L. Adamson, S. Wise, & T. B. Brazelton (1978). The infant's response to entrapment between contradictory messages in face-to-face interaction. *Journal of the American Academy of Child Psychiatry*, 17, 1–13.
- Zahavi, D. (2004). The embodied self-awareness of the infant. A challenge to the theory-theory of mind? In D. Zahavi, J. Parnas, & T. Grünbaum (Eds.), *Interdisciplinary perspectives on self-consciousness*. Amsterdam: John Benjamins.

CHAPTER 3

The embodied self-awareness of the infant

A challenge to the theory-theory of mind?

Dan Zahavi

University of Copenhagen

The aim of the following contribution is to discuss whether recent findings in developmental psychology, findings concerning infantile self- and other-experience, might challenge a view held by advocates of the theory-theory of mind, namely the view that both self-awareness and intersubjectivity presuppose a theory of mind.

1. Theory of mind

The term “theory of mind” was originally introduced by Premack and Woodruff in a seminal paper on intentionality in primates:

In saying that an individual has a theory of mind, we mean that the individual imputes mental states to himself and to others (either to conspecifics or to other species as well). A system of inferences of this kind is properly viewed as a theory, first, because such states are not directly observable, and second, because the system can be used to make predictions, specifically about the behavior of other organisms. (Premack & Woodruff 1978:515)

The phrase “theory of mind” was consequently used as shorthand for our ability to attribute mental states – such as intentions, beliefs, and desires – to self and others and to interpret, predict, and explain behavior in terms of mental states.¹ However, although Premack and Woodruff took it for granted that it was the possession and use of a theory that gave the individual the capacity to attribute mental states, the contemporary debate is split on this issue. On one side, we have the theory-theory of mind, and on the other the simulation theory of mind. The theory-theorists claim that the ability to explain and

predict behavior is underpinned by a folk-psychological theory dealing with the structure and functioning of the mind. We attribute beliefs to others by deploying theoretical knowledge. There is, however, disagreement among the theory-theorists about whether the theory in question is innate and modularized (Carruthers, Baron-Cohen), or whether it is acquired in the same way as scientific theories are acquired (Gopnik, Meltzoff). Most claim that there is some innate basis, but as Gopnik points out, it is necessary to distinguish between modularity nativism and starting-state nativism. The *theory-formation* theory, which takes the child to be a little scientist, who is constructing and revising theories in the light of incoming data, can accept a certain nativism, but such initial structures are taken to be defeasible. They can be changed and will be changed by new evidence (Gopnik 1996:171). Thus, for the theory-formation theory, there is a striking similarity between the acquisition of scientific knowledge and the child's increasing ability to adopt the intentional stance and mind-read, i.e., his or her ability to interpret behavior in terms of an agent's mental state. The same cognitive processes are responsible for scientific progress and for the development of a child's understanding of the mind (Gopnik 1996: 169). In contrast, the *modularists* claim that the core of the folk-psychological theory is hardwired. As they point out, if the theory were merely the product of scientific investigation, why is it culturally universal and why do all children reach the same theory at the same age (Carruthers 1996a:23). According to the modularists, the theory is forged by evolution and innately given, and although it might need experience as a trigger, the theory of mind module will not be modified by experience.

Whereas the theory-theorists claim that we employ a theory about the psychological when we predict and explain the behavior of others, the simulationists claim that we possess no such theory, or at least none complete enough to underpin all our competence with psychological notions (Heal 1996:75). Whereas the theory-theorists make use of what Gordon has called a cold methodology and argue that our understanding of others chiefly engages intellectual processes, moving by inference from one belief to the other, the simulationists employ a hot methodology and argue that our understanding of others exploits our own motivational and emotional resources (Gordon 1996:11). Thus, according to the simulationists, what lies at the root of our mind-reading abilities is not any sort of theory, but rather an ability to project ourselves imaginatively into another person's perspective, simulating his or her mental activity with our own.

In the following, my focus will be on the theory-theory.² What precisely do the theory-theorists mean by theory? We have already seen that the view

differs. Some take the theory of mind to be a theory in a very literal sense and compare it to a scientific theory.³ Others regard it in a more extended sense and compare it to a set of rules of symbol manipulation instantiated in an innate module. Some take the theory in question to be explicit, to be something the agent is conscious of, others consider it to be more or less implicit and tacit, and to be something that operates on a subpersonal level.

This issue is of crucial importance not only in order to understand what is actually at stake, but also in order to properly evaluate the debate between the theory-theory and the simulation-theory. Unfortunately, however, not everybody has been careful to spell out what precisely they mean when they say that our understanding of the mental is underpinned by a theory. Generally speaking, however, many theory-theorists have tended to construe theory in a rather loose sense, in order to increase the plausibility of their own position, but the danger they thereby run is to become vulnerable to what is known as the “promiscuity objection” (cf. Blackburn 1995). In the end, the notion of theory becomes vacuous, and everything turns out to be theoretical, including cooking, gardening, and fishing. In order to avoid this, some theory-theorists have simply bitten the bullet, and have accepted a strong definition of theory that entails much more than simply some kind of semantic holism.

The theory-theory claims that mastery of mental concepts is constituted by knowledge of a psychological theory. More specifically, our understanding of mental notions depends upon our knowledge of the positions that these notions occupy within the theory. Thus, the notions are thought to receive their sense from the theory in which they are embedded, rather than through some ostensive definition or direct acquaintance. This is probably one of the most characteristic features of the theory-theory: It denies that our reference to mental states such as beliefs and desires is based on any direct experience of such mental states, and instead argues that the concepts in question are theoretical postulates that have been developed through a process of abstract theorizing. To put it differently, since the theory-theorists consider the attribution of mental states to be a question of an inference to best explanation and prediction of behavioral data, they have often taken mental states to be unobservable and theoretically postulated entities. Leslie vividly articulates such a view in the following passage:

One of the most important powers of the human mind is to conceive of and think about itself and other minds. Because the mental states of others (and indeed of ourselves) are completely hidden from the senses, they can only ever be inferred.
(Leslie 1987: 139; cf. Baron-Cohen 1995: xvii)

2. Theory-theory of self-awareness

According to many theory-theorists (Gopnik, Carruthers, Frith and Happé) – at least if one takes some of their explicit statements at face value – we come to know our own beliefs and occurrent mental states just like we come to know the beliefs and experiences of others. In both cases, the same cognitive mechanism is in use, in both cases we are dealing with a process of mind-reading, in both cases we are dealing with the application of a theory of mind. Thus, according to what might be labeled the theory-theory account of self-awareness, my access to my own mind depends on the same mechanisms that I use in attributing mental states to others. In both cases, the access, the understanding, and the knowledge are theory-mediated, and the mediating theory is the same for self and for other (Carruthers & Smith 1996: 3; Gopnik 1993: 3; Frith & Happé 1999: 7). Even though we seem to perceive our own mental states directly, this direct perception is an illusion.

The theory-theory predicts that there should be no difference in the development of our ability to attribute mental states to self and other, since the same cognitive mechanism is used in both cases. In other words, the individual's ability to mind-read should be equally good (or bad) regardless of whether the tasks concern his own mental states or the mental states of others.⁴ The existence of an extensive parallelism would consequently provide empirical support for the theory-theory. Does such a parallelism in fact exist? In order to investigate the matter, a whole battery of tests has been used.⁵ Let me focus on the most well known tests: the *false-belief* tasks.

The two most frequently used false-belief tasks are the location change and the content change tasks. The *Sally-Anne task*, a location change task, is set up in the following manner. The child is confronted with two dolls, Sally and Anne. Sally has a box and Anne has a basket. Sally puts a marble into her box and then goes for a walk. While she is away, Anne takes the marble from the box, and puts it into her own basket. Sally then returns. She wants to play with her marble, but where will she look for it? When four-year old children (and older) are confronted with the question, they typically say that she will look inside her box, since that is where she *falsely believes* it to be hidden. Younger children, however, often point to the basket, indicating that they think that Sally will look for the marble where it really is. They apparently fail to understand that other persons' beliefs may be false (Frith & Happé 1999: 3–4).

In the *Smarties-task*, a content change task, children are shown a candy box. Based on its appearance, children first believe that the box contains sweets, but the box is then opened and is shown to contain pencils. The box is then closed

again, and the children are asked what other children, who have not yet seen inside the box, will think it contains. The average four-year-old answers that other children will think it contains candy, whereas younger children answer pencils. Once again, the result seems to demonstrate that very young children are unable to comprehend that other persons might have false beliefs.

Why is there this interest in children's ability to succeed on false-belief tasks? Because, in order for a child to ascribe false beliefs to others (and to himself), he must supposedly be able to understand that our beliefs might differ from reality. In order to make sense of Sally's behavior, for instance, the young child has to understand that Sally is acting not on the basis of what is actually the case, but on the basis of a false belief about what is the case. Thus, the young child must be able to understand the difference between reality and our beliefs about reality. She must have beliefs about beliefs; she must be in possession of a theory of mind.

In order to test the existence of a parallelism in the attribution of mental states to self and to other, an ingenious variation of the *Smarties-task* was devised. Children were presented with the deceptive candy box that was full of pencils. They were first asked the above-mentioned questions, and then the following question was added: "When you first saw the box, before we opened it, what did you think was inside it?" Somewhat surprisingly, one-half to two-thirds of the three-year-olds said that they had originally thought that it contained pencils. They apparently failed to remember their own past false beliefs. Thus, three-year-old children seem to have as much trouble understanding their own past false beliefs as they have in understanding the false beliefs of others (Gopnik 1993:6–8).

According to Gopnik, this finding reveals a striking parallelism between children's understanding of the psychological states of others and their understanding of their own immediate-past psychological states. But this parallelism does not support our commonsense intuition that the process of discovering our own mental states is fundamentally different from the process of discovering someone else's states. According to the commonsense view, there is phenomenological asymmetry between self and other. Whereas we can observe the other's behavior and have to infer his beliefs and desires, we have direct access to our own beliefs and desires and can simply report them. We need not infer their existence; we do not need any theoretical model at all. But for Gopnik, the existence of the parallelism challenges this view. In fact, when children can report and understand the psychological states of others, they can report having had those states themselves, and when they cannot report and understand the psychological states of others, they report that they have not had those states

themselves (Gopnik 1993:9). In short, there is little evidence that mental states are attributed to self before they are attributed to others, and vice versa.

If our acquisition of beliefs about our own mental states parallels our acquisition of beliefs about the mental states of others, and if the epistemic source is fundamentally the same in both cases, why do we normally tend to believe that there is such a big difference between the two? The explanation offered by both Gopnik and Carruthers is that we have become experts on reading our own minds, and after having reached a certain expertise we tend to see things at once, although what we see is actually the result of a complex theoretical process. We draw on an accumulated theoretical knowledge, but our expertise makes us unaware of the inferential processes and makes us believe that our experience is immediate and non-inferential. In other words, self-knowledge or self-consciousness can be thought of in analogy with the theory-laden perception of theoretical entities in science. Just as a diagnostician can sometimes see a cancer in the blur of an x-ray picture, so, too, each of us can sometimes see that we are in a state accorded such-and-such a role by folk-psychological theory (Gopnik 1993: 11; Carruthers 1996a: 26; 1996b: 259–260).

Gopnik argues that developmental evidence confirms the existence of a crucial parallelism in the attribution of mental states to others and in the attribution of mental states to self. However, the most significant of the findings presented by Gopnik demonstrates the existence of a parallelism in the attribution of *current* false beliefs to others and in the attribution of *past* false beliefs to self, but it is rather unclear why these findings – puzzling and interesting as they are – should warrant the kind of sweeping claim made by Gopnik. Apparently, however, Gopnik's idea is that unless you are able to appreciate that it is possible to have mistaken beliefs, you cannot understand what it means to have beliefs or intentional states at all (Gopnik 1993:6). It is certainly reasonable to assume that if a child can understand what a false belief is, then she can also understand what a belief is. But is it also reasonable to conclude that unless a child can understand false beliefs she cannot understand beliefs? Certainly, if we are talking about a full-fledged theoretical understanding of beliefs, i.e., of an actual theory of beliefs. Such a theory must involve an understanding and explanation of the possibility of error. If it did not, we would likely say that it was not really a theory of beliefs, or at best, that it was a very inadequate theory of beliefs. However, it is hardly surprising that we have high requirements for what a theory should entail. The question is whether it is appropriate to apply the same strong requirements to a young child, and to claim that the young child doesn't experience himself as having intentional states unless he

masters a theory of mind, unless he is capable of attributing false-beliefs to self and others.

In a more recent article, Nichols and Stich (2002) launched a rather damning attack on the theory-theory account of self-awareness. As they pointed out, there are three ways to interpret the theory. First, it could be taken to involve the claim that the only information we have about our own mental states is the kind of evidence that others are also in possession of. In this sense, knowledge of self and knowledge of others would be completely analogous. However, this is a form of pure behaviorism that is hopelessly implausible. Second, the theory could be taken to involve the concession that my access to my own mental states is based on information that is not available in the case of my access to the mental states of others. The problem, however, is that the theory never spells out what precisely this information is. Gopnik refers to first-person psychological experience as “the Cartesian buzz” (Gopnik 1993: 11), but as Nichols and Stich point out, this is not a very illuminating answer. Finally, the theory-theory account of self-awareness might argue that the additional information that is available in my own case is information about my own mental states. But if this information is available to me from the outset, there is no reason to introduce and involve any theory of mind mechanism (Nichols & Stich 2002: 12).

I agree with this criticism, but in the following, I will consider a somewhat different challenge to the theory-theory of mind. As we have just seen, the theory-theory claims that self-awareness – in the sense of having access to or being acquainted with one’s own mind – is theoretical in nature and that it presupposes a theory of mind. According to the standard view, however, children only gain possession of a theory of mind when they are around four-years old.⁶ It is only at that age that they can pass the classical theory of mind tasks, such as the false-belief task or the appearance-reality task. And as both Baron-Cohen and Frith and Happé have argued, one can test the presence of self-awareness using these classical tests (Frith & Happé 1999: 5; Baron-Cohen 1989: 591). As Baron-Cohen for instance puts it, since the ability to understand the appearance-reality distinction involves the ability to attribute mental states to oneself, a failure to pass the task suggests a lack of self-awareness (Baron-Cohen 1989: 596).⁷

In order fully to understand the theory-theory perspective on self-consciousness, it might be useful to recall that theory-theorists are committed to some version of the higher-order account of consciousness. This commitment is rarely spelled out, but it is crucial to their overall line of argumentation. Carruthers is a theorist who has not only undertaken the trouble of actually spelling out the link between the theory-theory of mind, the higher-order

thought theory, and the issues of self-awareness and phenomenal consciousness, but who has also done so with exemplary lucidity and characteristic bluntness. Carruthers takes conscious mental states, that is, mental states with a distinctive subjective feel to them, mental states that it feels like something to be the subject of, to be mental states of which the subject is aware, and he consequently argues that conscious mental states require self-awareness, or to put it differently, he argues that self-awareness is a conceptually necessary condition for there to be phenomenal consciousness (Carruthers 1996c: 155). Carruthers considers the self-awareness in question to be a type of higher-order thinking, and he therefore argues that a creature must be able to think about and hence conceptualize its own mental states if these states are to feel like anything to the organism. Thus, to have a phenomenally conscious perception of a surface as green, the creature must entertain the higher-order thought “I am perceiving a green surface.” Since mental concepts get their significance from being embedded in a folk-psychological theory of the structure and functioning of the mind, what this ultimately means is that only creatures in possession of a theory of mind are capable of enjoying conscious experiences (Carruthers 1996c: 158, 2000: 194). As he puts it:

[I]n order to think about your own thoughts, or your own experiences, you have to possess the *concepts* of thought and experience. And these get their life and significance from being embedded in a folk-psychological theory of the structure and functioning of the mind. So in the case of any creature to whom it is implausible to attribute a theory of mind – and I assume that this includes most animals and young infants – it will be equally implausible to suppose that they engage in conscious thinking. [...] If animals (or most animals) lack higher-order thoughts, then by the same token they will lack conscious experiences. For there will be just as little reason to believe that they are capable of thinking about their own experiences, as such. If true, this conclusion may have profound implications for our moral attitudes towards animals and animal suffering. (Carruthers 1996c: 221; cf. 2000: 194)

Carruthers consequently holds the view that animals (and infants under the age of three) lack phenomenal consciousness, lack a dimension of subjectivity. In his view, they are blind to the existence of their own mental states; there is in fact nothing it is like for them to feel pain or pleasure (Carruthers 1998: 216, 2000: 203). Carruthers concedes that most of us believe that it must be like something to be a young infant, a cat, or a camel, and that the experiences of these creatures have subjective feels to them, but he considers this common-sense belief to be quite groundless (Carruthers 1996: 223).⁸

If a theory of mind is required for self-awareness, any creature that lacks such a theory will also lack self-awareness. Is it true, however, that infants lack self-awareness during the first 3–4 years of life? To suggest that an infant only becomes self-aware when he is in possession of a theory of mind, or to mention some other traditional candidates, when he masters the use of the first-person pronoun, or when he is able to recognize himself in the mirror, in my view, is to operate with an unacceptably narrow definition of self-awareness.⁹ It is also a suggestion that a number of prominent developmental psychologists have criticized. However, if it could be shown that infants are in possession of self-awareness before they acquire a theory of mind, the theory-theory would be in trouble.

3. Developmental counter-evidence

Let us take a closer look at some of the empirical findings that are discussed in the work of Stern, Neisser, Butterworth, and Rochat. These developmental psychologists argue that the infant is in possession of self-experience from birth, and they all reject the view, originally defended by Piaget, according to which the infant initially lives in a kind of a dualistic fusion where there is as yet no distinction between self, world, and other (Piaget & Inhelder 1969:22). Thus, according to this once widely held view, the infant was initially supposed to exist in a “state of undifferentiation, of fusion with mother, in which the ‘I’ is not yet differentiated from the ‘not-I’ and in which inside and outside are only gradually coming to be sensed as different” (Mahler, Pine, & Bergman 1975:44).

If we start with Stern, he argues that theory and language *transform* and *articulate* the infant’s experience of self and other; they do not constitute it. Already from birth onward, the infant gains possession of different pre-reflective and pre-linguistic “senses of self”. Stern concedes that the sense of self initially available to the infant is basic, but he lists four types of experiences that are present at around three months of age. There is *self-agency*, that is, the sense of authorship of one’s own actions; there is *self-coherence*, that is, the sense of being an integrated, non-fragmented whole; there is *self-affectivity*, that is, the experience of subjective feelings; and finally there is *self-history*, that is, the having of a sense of endurance, of being in continuity with one’s own past (Stern 1985:71). These four experiences are all basic types of self-experience and, according to Stern, they are not merely cognitive constructs, but rather lived, existential counterparts to the objectifiable, verbalizable self.

It would lead us too far astray to discuss Stern's analyses of all four types in detail, but let me focus on his account of self-agency or authorship of actions. How does a child distinguish between her own movements/actions and the movements/actions of others, and what enables her to experience *herself* as an agent? Stern distinguishes between two *experiential* invariants: 1) The sense of volition that precedes a motor act, and 2) the proprioceptive feedback that does or does not occur during the act (Stern 1985:76). The child typically encounters three different types of action: self-willed action of self, other-willed action of other, and other-willed action of self. Further, the child is able to distinguish between the three precisely because of the presence or absence of invariants 1 and 2. If the experience of the action contains both volition and proprioceptive feedback, we are dealing with a self-willed action of self. If neither is present, we have an other-willed action of other. And if the proprioceptive feedback is present, but the experience of volition is absent (as in the case where the mother is moving the hand of the infant), we have an other-willed action of self.

Just like Stern, Neisser, Butterworth, and Rochat also reject the view that self-awareness has a late developmental onset. In a well-known article from 1988, Neisser distinguished five different selves: the ecological self, the interpersonal self, the extended self, the private self, and the conceptual self (Neisser 1988:35). The most basic and primitive of these is the ecological self, that is, the individual understood as an active agent in the immediate environment. When and how are we aware of the ecological self? According to Neisser, this occurs whenever we perceive. Following Gibson, Neisser takes perception to involve information about the relation between the perceiver and the environment. In this sense, all perception involves a kind of self-sensitivity; all perception involves a co-perception of self and of environment (cf. Gibson 1986:126). As perceivers, we are embedded and embodied agents. We see with mobile eyes that are set in a head that can turn and that is attached to a body that can move from place to place; in this sense a stationary point of view is only the limiting case of a mobile point of view (Gibson 1986:53, 205). But every movement of the perceiver produces a systematic flow pattern in the visual field, which provides us with awareness of our own movements and postures. Thus, proprioception (or kinaesthesia) is richly intermodal; it is neither attached to a unique sense-organ, nor is it to be identified with a specific body sense. Rather, it is a mechanism of self-sensitivity, common to all perceptual systems. It can be obtained through vision or audition, as well as through the muscles and joints.

Employing the Gibsonian notion of affordance, Neisser writes that any given situation affords some actions and not others. We see at a glance whether objects are within reach, whether doors are wide enough to walk through, or

chairs are the right height to sit on. Moreover, this perception is “body-scaled”, that is, the distance that matters is not measured in centimeters, but in relation to our own bodily dimensions and capabilities (Neisser 1993:8). For instance, a young infant (a few weeks old) can discriminate between objects that are within his reach and objects that are outside his reach. The infant is far less inclined to reach out for an object that is outside his reach. But, of course, for the infant to be able to make this distinction, he must be aware of the position of the object in relation to *himself*. That is, the infant has to be in possession of *self-specifying information*. Even very young infants pick up the information that specifies the ecological self. They respond to the optical flow, discriminate between themselves and other objects, and easily distinguish their own actions and their immediate consequences from events of other kinds. They perceive themselves (among other things), they perceive where they are, how they are moving, what they are doing, and whether a given action is their own or not. These achievements appear already in the first weeks and months of life, and, according to both Butterworth and Neisser, they testify to the existence of a primitive and irreducible form of self-awareness (Neisser 1993:4; Butterworth 2000:24).

According to Rochat, newborn infants (24 hours old) can discriminate between double touch stimulation combined with proprioception and single touch of exogenous origin. All healthy infants have an innate rooting response. When the corner of an infant’s mouth is touched, the infant turns her head and opens her mouth toward the stimulation. By recording the frequency of rooting in response to either external tactile stimulation or tactile self-stimulation, it was discovered that newborns showed rooting responses almost three times more frequently in response to the external stimulus. Rochat thus concludes that even newborns can pick up the intermodal invariants that specify self-versus nonself-stimulation, and that they thereby have the ability to develop an early sense of self (Rochat 2001:40–41). Infants are in possession of proprioceptive information from birth and as Rochat argues, proprioception is “the modality of the self par excellence” (Rochat 2001:35). Thus, long before they are able to pass any mirror self-recognition tasks, not to speak of any false-belief tasks, infants have a sense of their own bodies as organized and environmentally embedded entities. They have an early sense of their own bodies, and hence an early perceptually-based sense of themselves (Rochat 2001:41). Following in the footsteps of Neisser and Gibson, Rochat calls this early sense of self the infant’s ecological self (Rochat 2001:30–31).

For Rochat, the ecological self is clearly a bodily self, and he argues that the infant’s self-experience is initially a matter of the infant’s experience of his own

embodied self. It is through their early body exploration that infants specify themselves as differentiated agents in the environment, eventually developing an explicit awareness of themselves. More precisely, infants have an inborn inclination to investigate their own bodies. This inclination forms the cradle of self-perception and constitutes the developmental origin of self-knowledge (Rochat 2001: 29, 39, 74).

Around the age of fifteen to eighteen months, the child becomes able to perform symbolic actions, and it acquires some linguistic competence. That the child becomes able to assume a detached perspective on itself can be seen for instance from its behavior before a mirror. Prior to this age, the child presumably does not realize that it sees itself in the mirror. If one marks the face of a child with rouge without her knowledge and she subsequently looks in a mirror, a younger child will point to the mirror and not to herself. But after the age of eighteen months, the child will touch the rouge on her own face. Since the confrontation with the mirror motivates a *self-directed* behavior, it is assumed that the child now recognizes what she sees in the mirror as her own reflection (Lewis & Brooks-Gunn 1979: 33–46; Stern 1985: 165).

However, although this recognition testifies to the existence of self-awareness, its absence certainly does not imply a lack of self-awareness. Not only is the recognition of one's own reflection by no means a primitive and basic type of self-awareness, on the contrary, we are dealing with a rather sophisticated type of representationally mediated self-identification, where the self-awareness in question takes place across distance and separation. We identify "that other" as ourselves. Moreover, the child would not be able to perform this identification, which presumably takes place through the perfect match between his own bodily movements and the movements of the mirror image, if he were not already aware of his *own* bodily movements. In short, in order to recognize oneself in the mirror, one must already be in possession of bodily self-awareness.

All of these authors are pointing to a dimension of bodily self-experience that is in place long before the infant is capable of solving any theory of mind tasks. Insofar as the theory-theory wants to uphold the view that all self-awareness is theoretically mediated, it is confronted with a serious problem. Let us not forget, however, that the theory-theory of mind defends a double thesis. It is not only claiming that self-awareness is theoretically mediated, it is also claiming that intersubjectivity is theoretically mediated. After all, the whole idea is that any reference to minded beings (be it to oneself or to others) involves a process of mind-reading, involves an application of a theory of mind. Given this situation, it is natural to ask whether the theory-theory treatment

of intersubjectivity might also be beset with related empirical and conceptual difficulties.

4. Embodiment and intersubjectivity

Infants are in possession of a form of bodily self-awareness long before they are in possession of a theory of mind, long before the infant is able to pass any theory of mind tasks. Moreover, they are certainly also capable of social interaction at this early stage. Whereas we, in adult life, occasionally make inferential attributions of mental states to other people, such attributions cannot be considered the basis of the smooth and immediate interpersonal interaction – often called primary intersubjectivity – found in young infants (Trevvarthen 1979). In some respects, the period between two and six months might be classified as the most social period in one's life. The social smile is already in place, and the child has a clear preference for perceiving other subjects rather than inanimate objects (Stern 1985:63, 72; Spitz 1983:98–124). Although an infant initially has very little command over her own locomotion, she has an almost fully developed control over her eye-movements, and can function as a social partner through her gaze. By controlling her own direction of gaze, she can regulate the level and amount of social stimulation. And through gaze behaviors, such as averting her gaze, shutting her eyes, staring past, becoming glassy-eyed, etc., to a large extent she can initiate, maintain, terminate, and avoid social contact (Stern 1985:21).

Two to three month-old infants will engage in “protoconversations” with other people by smiling and vocalizing, and will demonstrate a capacity to vary the timing and intensity of communication with their partners. The purpose of this early interaction seems to be the interaction itself, with the participants affectively resonating to one another (Fivaz et al. 2004). When a mother mirrors the infant's affect, the infant will reciprocate and show sensitivity to the affective mirroring of the mother. In fact, infants clearly expect people to communicate reciprocally with them in face-to-face interactions, and to work actively with them in order to sustain and regulate the interaction. If the mother is asked to remain immobile and unresponsive, the infant will react by ceasing to smile, and will exhibit distress and attempt to regain her participation.

Infants are clearly reacting differently to mere objects and other subjects from the very start. Whereas objects are simply toys to be looked at and manipulated, the faces, voices, and bodily movements of other people are treated as special social parameters (Legerstee 1999:217, 220–221). Infants are also able

to interpret the bodily movements of others as goal-directed and intentional, in short, they have the capacity to perceive others as agents. And there is nothing inferential about this early social interaction; rather, it is a form of intersubjectivity based on the infant's intuitive grasp of the expressive gestures of other individuals.

Around the age of nine months, a change occurs, insofar as the infant starts to realize that it can share experiences of the world with others. This change in the infant's experience of self and other is evinced from the infant's attempt to share joint attention, intentions, and affective states (Stern 1985: 128). As Rochat writes:

Research shows that by nine months infants begin to treat and understand others as "intentional agents", somehow explicitly recognizing that like themselves, people plan and are deliberate in their actions. So, for example, infants will start sharing their attention toward objects with others, looking up toward them to check if they are equally engaged. They will start to refer to other people socially, and in particular to take into consideration the emotional expression of others while planning actions or trying to understand a novel situation in the environment. (Rochat 2001: 185)

Infants of nine months can follow the eye-gaze or pointing finger of another person, and when they do so, they often look back at the person and appear to use the feedback from his or her face to confirm that they have in fact reached the right target. In other words, they seek to validate whether joint *attention* has been achieved. Similarly, they might show objects to others, often looking to the other person's eyes, to check whether he or she is attending. As for the sharing of *intentions*, it is most obvious in protolinguistic requests for help. Such requests suggest that the infant apprehends the other as someone who can comprehend and satisfy her own intentions. Similarly, they might respond to simple verbal requests by others, or shake their heads to express refusal. Thus, intentions have become shareable experiences (Stern 1985: 129–131). Finally, the sharing of *affection*, or interaffectivity, which is presumably the first and most basic form of subjective sharing, can also be witnessed. If an infant is placed in a situation that is bound to generate uncertainty, for instance, by being approached by a new, unusual, and highly stimulating object, such as a bleeping and flashing toy, he will look toward his mother for her emotional reaction, essentially to see what he should feel in order to help resolve his own uncertainty. If the mother shows pleasure by smiling, the infant will continue his exploration; if she shows fear, the infant will turn back from the object and perhaps become upset (Stern 1985: 132). A vivid example of this is the famous

“visual cliff” experiment. Infants aged twelve months are placed on one side of a “visual cliff”, i.e., an apparent sudden drop beneath a transparent surface. On the other side of the cliff, the infant’s mother and an attractive toy are placed. When the infant notices the drop-off, she will typically look spontaneously at her mother’s face. If the mother poses a happy face, most infants will cross to the deep side; if the mother poses a fearful expression, the infants will freeze or even actively retreat. It is noteworthy that the mother’s mere presence is not enough, rather her emotional reaction, as perceived through her expressions and behavior, has a decisive influence (Hobson 1991:47). In other words, the infant appears to recognize that another person’s expression has meaning with reference to an environment common to both of them. The gestures and utterances of the caretaker are perceived as being both emotionally expressive and as being directed to something in the infant’s world (Hobson 1993:38, 140–141). Thus, Hobson concludes that infants

...have direct perception of and natural engagement with person-related meanings that are apprehended in the expressions and behaviour of other persons. It is only gradually, and with considerable input from adults, that they eventually come to conceive of ‘bodies’ on the one hand, and ‘minds’ on the other. (Hobson 1993:117)

Are embodied self-experience and the experience of others linked? Some philosophers have argued that unless self-experience is embodied, intersubjectivity is neither possible nor comprehensible. To put it differently, if we adopt what McCulloch has recently called a behavior-rejecting mentalism (McCulloch 2003:94), i.e., if we deny that embodiment and bodily behavior have any essential role to play in experience and cognition, if we deny that embodiment and environmental embedding are essential to having a mind, we will have a hard time escaping solipsism.

What does the argument look like? If my own self-experience, in the first instance, is of a purely mental nature, if my embodiment does not figure in my self-acquaintance from the very start, we need to understand how I will ever be inclined to attribute selfhood to others. Why should I even so much as think that there are other selves? Had subjectivity been an exclusive first-person phenomenon, were it only present in the form of an immediate and unique inwardness, I would only know one case of it – my own – and would have had no reason to ascribe it to others, and to recognize other bodies as embodied subjects. To quote Merleau-Ponty and Davidson:

If the sole experience of the subject is the one which I gain by coinciding with it, if the mind, by definition, eludes “the outside spectator” and can be recog-

nized only from within, my *cogito* is necessarily unique, and cannot be “shared in” by another. Perhaps we can say that it is “transferable” to others. But then how could such a transfer ever be brought about? What spectacle can ever validly induce me to posit outside myself that mode of existence the whole significance of which demands that it be grasped from within? Unless I learn within myself to recognize the junction of the *for itself* and the *in itself*, none of those mechanisms called other bodies will ever be able to come to life; unless I have an exterior others have no interior. (Merleau-Ponty 1945:427–428)

If the mental states of others are known only through their behavioral and other outward manifestation, while this is not true of our own mental states, why should we think our own mental states are anything like those of others? (Davidson 2001:207)

The basic problem is as follows: If my body does not figure essentially in my self-ascription of (some) mental terms and if my ascription of mental terms to others is essentially based on their bodily behavior and expression, what should then guarantee that we are in fact applying the same concepts to ourselves and to others? The different uses of the concepts threaten the unity of their meaning (cf. Avramides 2001:135, 224). The proper way to respond to this skeptical challenge is by abandoning the radical divide between the subject’s mind and body, and one way to do so is by appealing to the notion of *action*. Action – at least according to one venerable philosophical tradition – joins mind and body, or more precisely, action is prior to the artificial division between mind and body.

It could be argued, of course, that any account of the mind has to take subjectivity and the first-person perspective seriously, and that a focus on behavior and action will consequently lose what is essential to the mind. However, as Avramides points out, this worry is simply misguided. There is nothing reductive in the reference to action, since subjectivity figures centrally in the concept. Action is the action of subjects; it is the action of minded individuals (Avramides 2001:286). We must respect the difference between the first-person and the second- and third-person perspectives and we should recognize the difference between self- and other-ascription. But too much focus on this difference or asymmetry can lead to the mistaken view that only my own experiences are given to me, and that the behavior of the other shields his experiences from me and makes their very existence hypothetical (Avramides 2001:187).

Merleau-Ponty has been very explicit in linking the issues of embodiment and intersubjectivity. In his view, subjectivity is essentially embodied. To exist embodied is, however, neither to exist as pure subject nor as pure object, but to exist in a way that transcends both alternatives. It does not entail a

loss of self-awareness; on the contrary, self-awareness is intrinsically embodied self-awareness, but it does entail a loss or perhaps rather a release from transparency and purity, thereby permitting intersubjectivity. As Merleau-Ponty writes: “The other can be evident to me because I am not transparent for myself, and because my subjectivity draws its body in its wake” (Merleau-Ponty 1945:405). To put it differently, since intersubjectivity is a fact, there must exist a bridge between my self-acquaintance and my acquaintance with others; my experience of my own subjectivity must contain an anticipation of the other, must contain the seeds of alterity (Merleau-Ponty 1945:400–401, 405, 511). If I am to recognize other bodies as embodied foreign subjects, I have to be in possession of something that will allow me to do so. But as Merleau-Ponty points out, when I experience an other and when I experience myself, there is in fact a common denominator. In both cases, I am dealing with *embodiment*, and one of the features of my embodied subjectivity is that it per definition comprises an *exteriority*. When I go for a walk, or write a letter, or play ball – to use Strawson’s examples (Strawson 1959:111) – I am experiencing myself, but in a way that anticipates the manner in which I would experience an other, and an other would experience me. This is not to say that a focus on embodiment and action eradicates the difference between self-ascription and other-ascription, between a first-person perspective and a second-person perspective, but it conceives of the difference in such a manner that their relationship becomes more intelligible. Thus, Merleau-Ponty can describe embodied self-awareness as a presentiment of the other and the experience of the other as an echo of one’s own bodily constitution. In short, it is because I am not a pure interiority, but an embodied being that lives outside itself, that transcends itself, that I am capable of encountering and understanding others who exist in the same way (Merleau-Ponty 1960:213, 215, 221, 1964:74).

The idea is not to reduce consciousness as such to intentional behavior. Rather, the idea is simply that bodily behavior, expression, and action are essential to (and not merely contingent vehicles of) some basic forms of consciousness. Mental states do not simply serve to explain behavior; rather some mental states are directly apprehended in the bodily expressions of people whose mental states they are. Or as Hobson also puts it: “We perceive bodies and bodily expressions, but we do so in such a way that we perceive and react to the mental life that those physical forms express” (Hobson 2002:248; cf. 1993:184). More generally, there seems to be something very problematic about claiming that intersubjective understanding is a two-stage process of which the first stage is the perception of meaningless behavior, and the second an intellectually-based attribution of psychological meaning. On the contrary, in the face-to-face en-

counter, we are neither confronted with a mere body, nor with a hidden psyche, but with a unified whole. When I see another's face, I *see* it as friendly or angry, etc., that is, the very face expresses these emotions. To quote Wittgenstein:

We do not see facial contortions and *make the inference* that he is feeling joy, grief, boredom. We describe a face immediately as sad, radiant, bored, even when we are unable to give any other description of the features.

(Wittgenstein 1980:§570)

In general I do not surmise fear in him – I *see* it. I do not feel that I am deducing the probable existence of something inside from something outside; rather it is as if the human face were in a way translucent and that I were seeing it not in reflected light but rather in its own.

(Wittgenstein 1980:§170)

A similar view has been advocated by both Merleau-Ponty and Scheler, who argue that the affective and emotional experiences of others are given for us *in* expressive phenomena. Anger, shame, hate, and love are not only qualities of subjective experience, but also types of behavior or styles of conduct, which are visible from the outside (Merleau-Ponty 1964: 52–53; Scheler 1973: 254).

This does not rule out that some mental states are covert, of course, but not all mental states can lack an essential link to behavior, if intersubjectivity is at all to get off the ground.¹⁰

Our experience and understanding of others are not infallible, but there is a decisive difference between our everyday uncertainty about what precisely others might be thinking about, and the nightmare vision of the solipsist. Although we might be uncertain about the specific beliefs and intentions of others, this uncertainty does not make us question their very existence. In fact, as Merleau-Ponty points out, our relation to others is deeper than any specific uncertainty we might have regarding them (Merleau-Ponty 1945: 415).

5. Phenomenological misgivings

There are good reasons (philosophical as well as empirical) for maintaining that body-awareness constitutes genuine self-experience. Unfortunately, however, and contrary to expectations, the accounts offered by Rochat, Butterworth, Neisser, and Stern are not always sufficiently clear on this.

In the introduction to his book *The Infant's World*, Rochat suggests that there are three fundamentally different and contrasted classes of experiences: the experience of self, of objects, and of other people (Rochat 2001: 27). I wholeheartedly agree with this division, which very much fits the received view

in phenomenology. Unfortunately, however, Rochat does not really respect his own division. He very soon starts to talk of the body as an object of exploration, and speaks of self-perception as a question of differentiating one's own body from *other* objects in the environment (Rochat 2001: 34, 37).

As for Butterworth, he has argued that proprioceptive self-awareness should be distinguished developmentally from higher-order consciousness or reflective self-awareness (where the self is the object of one's own cognition). However, Butterworth still speaks of the ecological self as being the object of one's own perception, and of primary consciousness as the state of being aware of the self as a thing or an object situated in the physical and social environment (Butterworth 2000: 19–20).

We find the very same take in Neisser, who repeatedly talks of the self as an object (Neisser 1988: 35, 39, 40). Although Neisser concedes that the ecological self is *per se* not an object of thought, he nevertheless considers it an object of perception (Neisser 1988: 41, 56).

If we return to Stern's multi-faceted analysis of the infant's self-experience, we come across a similar objectivistic strain. Stern occasionally makes it sound as if an infant's self-experience is a result of her ability to discriminate herself from others, and that this is merely an instance of her general ability to discriminate between different entities. He claims that the infant, far from being a *tabula rasa*, is predesigned to perceive the world in a highly structured fashion. Just as she very early is able to perceive and organize different stimuli into different natural categories, the infant has inborn capabilities that enable her to discriminate different gestalt constellations of stimuli in such a way that she can keep self and other separate. When the infant feels the caress of her mother, hears the voice of her father, and sees her own hand, she is not overwhelmed by a surge of unstructured sensations, but is able to distinguish between herself, her father, and her mother as three distinct entities. She recognizes that the behavior of different persons is differently structured; she distinguishes one agent from another (Stern 1983: 56–62), and is thereby ultimately able to discriminate the invariant structure that characterizes her own self-generated actions and experiences from the patterns belonging to the movement and actions of particular others (Stern 1985: 7, 65, 67).

These ways of describing and accounting for self-experience, however, are beset with a major problem. Even if an infant is able to distinguish between different entities in such a way that no confusion takes place, this does not answer the key question: How does the infant sense that one of these experiential configurations is *itself*? The answer given is not satisfactory. Although both Stern and Rochat acknowledge that the infant's (direct and immediate) experience

of proprioception and volition is of crucial importance (Rochat 2001: 89; Stern 1983: 65), they still make it sound as if self-awareness is a question of discriminating correctly between two types of objects. But this is to commit the mistake of equating self-experience with object-identification, as if the infant were first confronted with certain experiences that he then subsequently succeeded in identifying as his own.

Why is it problematic to conceive of the embodied self as an object, and of embodied self-awareness as a kind of object-awareness? To put it very simply, for something to be given as an object is for it to be given as something that transcends the merely subjective. For something to be given as an object of experience is for it to differ from the subjective experience itself. However, if this is so, if object-awareness always involves a kind of epistemic divide, if object-awareness always entails a distinction between the subject and the object of experience, object-awareness cannot help us understand self-awareness. After all, self-awareness is precisely supposed to acquaint us with our own subjectivity; it is not supposed merely to acquaint us with yet another object of experience. Perhaps it could be objected that there surely are cases where I am confronted with a certain object, and then recognize that the object in question is in fact myself. This is true of course, but this kind of objectified self-recognition can never constitute the most fundamental form of self-awareness. Why not? Because in order for me to recognize a certain object as myself, I need to hold something true of it that I already know to be true of myself. The only way to avoid an infinite regress is by accepting the existence of a non-objectifying self-acquaintance. To quote Sidney Shoemaker:

The reason one is not presented to oneself “as an object” in self-awareness is that self-awareness is not perceptual awareness, i.e., is not a sort of awareness in which objects are presented. It is awareness of facts unmediated by awareness of objects. But it is worth noting that if one were aware of oneself as an object in such cases (as one is in fact aware of oneself as an object when one sees oneself in a mirror), this would not help to explain one’s self-knowledge. For awareness that the presented object was ϕ would not tell one that one was oneself ϕ , unless one had identified the object as oneself; and one could not do this unless one already had some self-knowledge, namely the knowledge that one is the unique possessor of whatever set of properties of the presented object one took to show it to be oneself. Perceptual self-knowledge presupposes non-perceptual self-knowledge, so not all self-knowledge can be perceptual.

(Shoemaker 1984: 105)

This reasoning holds true even for self-knowledge obtained through introspection. That is, it will not do to claim that introspection is distinguished by the

fact that its object has a property that immediately identifies it as being me, since no other self could possibly have it, namely the property of being the private and exclusive object of precisely my introspection. This explanation will not do, since I will be unable to identify an introspected self as myself by the fact that it is introspectively observed by me, unless I know it is the object of *my* introspection, i.e., unless I know that it is in fact *me* that undertakes this introspection. This knowledge cannot itself be based on identification if one is to avoid an infinite regress (Shoemaker 1968:561–563).

To recapitulate, the problem with the account offered by Stern, Rochat, and Butterworth is that they conceive of the embodied self as an object, and of embodied self-awareness as a kind of object-awareness. What is the alternative? Perhaps, phenomenological writings on self-awareness and embodiment can get us further.

According to the phenomenologists, self-awareness should be construed very broadly. In contrast to what is claimed by the theory-theory, self-awareness is not something that only comes about the moment I construct a theory about the cause of my own behaviour, a theory that postulates the existence of mental states. Nor is it something that only comes about the moment one scrutinizes one's experiences attentively, not to speak of it being something that only comes about the moment one recognizes one's own mirror image, or refers to oneself using the first-person pronoun, or is in possession of identifying knowledge of one's own life story. Rather, literally all the major figures in phenomenology defend the view that the experiential dimension is as such characterized by a tacit self-awareness. They consequently take it to be legitimate to speak of self-awareness as soon as I am not simply conscious of a foreign object, but acquainted with the experience of the object as well, for in such a case my consciousness reveals itself to me. Thus, self-awareness is taken to be a question of having first-personal access to one's own consciousness; it is a question of the first-personal givenness or manifestation of experiential life. Most people are prepared to concede that there is necessarily something "it is like" for a subject to undergo a conscious experience (to taste ice cream, to feel joy, to remember a walk in the Alps). But insofar as there is something it is like for the subject to have the experience, the subject must in some way have access to and be acquainted with the experience. Moreover, although conscious experiences differ from one another – what it is like to smell crushed mint leaves is different from what it is like to see a sunset or to hear Lalo's *Symphonie Espagnole* – they also share certain features. One commonality is the quality of *mineness*, the fact that the experiences are characterized by first-personal givenness. That is, the experience is given (at least tacitly) as *my* experience, as an experi-

ence *I* am undergoing or living through. First-personal experience presents me with an immediate and non-observational access to myself. All of this suggests that we are dealing with a (minimal) form of self-awareness, that is, (phenomenal) consciousness is taken to entail a (weak) sense of self-awareness. To put it differently, self-awareness is taken to be a necessary condition for phenomenal consciousness. Unless a mental process is self-conscious, there will be nothing it is like to undergo the process, and it therefore cannot be a phenomenally conscious process.¹¹

The claim that there is a close link between consciousness and self-awareness is less exceptional than might be expected. In fact, it might be argued that such a claim is part of current orthodoxy, since higher-order theories typically take the difference between conscious and non-conscious mental states to rest upon the presence or absence of a relevant meta-mental state. To put it differently, (intransitive) consciousness has frequently been taken to be a question of the mind directing its intentional aim at its own states and operations. Thus, higher-order theories have typically taken self-directedness to be constitutive of (intransitive) consciousness.

But one might share the view that there is a close link between consciousness and self-awareness and still disagree about the nature of the link. And although the phenomenological take might superficially resemble the view of the higher-order theories, we are ultimately confronted with two radically divergent accounts. In contrast to the higher-order theories, the phenomenologists explicitly deny that the self-awareness that is present the moment I consciously experience something is to be understood in terms of some kind of reflection, or introspection, or higher-order monitoring. It does not involve an additional mental state, but is rather to be understood as an intrinsic feature of the primary experience.¹²

Of course, this is not to deny that there are also far more complex forms of self-awareness that are both theory- and language dependent and intersubjectively constituted, but the primitive self-awareness that is part and parcel of phenomenal consciousness is independent of such conceptual sophistication.

The phenomenological analysis of self-awareness complements the argumentation provided by the developmental psychologists, since it explicitly tackles an issue they largely remain silent about, namely the nature of experience and phenomenal consciousness. Most of the developmental evidence presented by Stern, Rochat, Neisser, and Butterworth is obviously behavioral in nature. However, in order for a creature to be in possession of self-awareness, it is not sufficient that the creature in question behaves in a certain way. It also has to be in possession of experiences, and it must behave as it does

because it has the experiences it has. To put it differently, any reasonable ascription of self-awareness cannot bypass a discussion of the relationship between the experiential dimension and self-awareness, but this is precisely what the phenomenological tradition can provide.

What about embodiment; how would the phenomenologists account for embodied self-awareness? Well, as Michel Henry once pointed out, a phenomenological clarification of the body must take its departure in the original givenness of the body (Henry 1965: 79). But how precisely is the body originally given? When I am watching a football match, I am normally not paying attention to the turn of my head when I follow the motions of the players, nor to the narrowing of my eyes when I attempt to discern the features of the goalkeeper. When I give up and reach for my binoculars, the movements of my hand remain outside the focus of my consciousness. When I am occupied with objects and directed at goals, my perceptual acts and their bodily roots are generally passed over in favor of the perceived, i.e., my body tends to efface itself on its way to its intentional goal. This is fortunate, because if we were aware of our bodily movements in the same way in which we are aware of objects, our bodies would make such high demands on our attention that it would interfere with our daily lives. However, when I execute movements without thinking about them, this is not necessarily because the movements are non-conscious, mechanical, or involuntary; rather, they might simply be part of my functioning intentionality, they might simply be immediately and pre-reflectively felt, as both Henry and Merleau-Ponty have argued (Henry 1965: 128; Merleau-Ponty 1945: 168). Thus, even if my movements might be absent as thematic intentional objects, this does not have to entail that they are experientially absent in any absolute sense.

Under normal circumstances, I do not need to perceive my arm visually in order to know where it is. If I wish to grasp the fork, I do not first have to search for the hand, since it is always with me. Whereas I can approach or move away from any object in the world, the body itself is always present as my very perspective on the world. That is, rather than being simply yet another perspectively given object, the body itself is, as Sartre and Merleau-Ponty points out, precisely that which allows me to perceive objects perspectively (Sartre 1976: 378; Merleau-Ponty 1945: 107). The body is present, not as a permanent perceptual object, but as myself. Originally, I do not have any consciousness *of* my body as an intentional object. I do not perceive it; *I am it*. Sartre even writes that the lived body is invisibly present, precisely because it is existentially lived rather than known (Sartre 1976: 372). This is also why Husserl repeatedly has emphasized how important it is to distinguish between *Leib* and *Körper*, that

is, between the pre-reflectively lived body, i.e., the body as an embodied first-person perspective, and the subsequent thematic experience *of* the body as an object (Husserl 1973:57).

In short, phenomenologists take pre-reflective body-awareness to be a question of how (embodied) consciousness is given to itself not as an *object*, but as a *subject*. Whereas Bermúdez has recently claimed that “somatic proprioception is a form of perception” that takes “the embodied self as its object” (Bermúdez 1998:132), the phenomenologists would argue that *primary* body-awareness is not a type of object-consciousness, is *not* a perception of the body as an object at all (cf. Gallagher 2003), but on the contrary a genuine form of self-experience.¹³

6. Conclusion

Let me by way of conclusion briefly summarize the results. According to the theory-theory, it is not only my understanding of other people’s mental states that involves a theory of mind. My access to my own mind also depends upon such a theory. In both cases, the same cognitive mechanisms are in use, in both cases we are dealing with a process of mind-reading, in both cases we are dealing with the application of a theory of mind. According to the standard view, however, children acquire a theory of mind at around the age of 4. It is only at this age they can pass the classical theory of mind tasks. Consequently, the theory-theory of mind argues that children lack proper self- and other-experience, during the first three to four years of life.

This view is faced with a number of both empirical and conceptual difficulties. Empirical findings strongly suggest that infants are in possession of a form of bodily self-awareness long before they are in possession of a theory of mind. Infants have an early sense of their own bodies; they are in possession of a dimension of bodily self-experience long before they are capable of solving any theory of mind tasks. Moreover, they are also capable of sophisticated social interaction that early.

At this stage, philosophy enters the picture. Phenomenologically inclined philosophers (including the later Wittgenstein) have not only argued that intersubjectivity presupposes embodiment and bodily self-experience, they have also – contrary to what has occasionally been claimed by developmental psychologists – denied that primary self-awareness is a type of *object*-consciousness. So to conclude, a recommended strategy is to combine the empirical findings of Stern, Neisser, Butterworth, and Rochat with the theoret-

ical considerations of Husserl, Sartre, Merleau-Ponty, and Wittgenstein. Jointly they constitute a serious challenge to the theory-theory of mind.¹⁴

Notes

1. The very choice of term is consequently quite revealing. It clearly indicates that psychological competence is taken to consist in the possession and use of a theory.
2. This (critical) focus on the theory-theory of mind should not be taken as an implicit endorsement of simulationism. There are problems with the simulation theory as well. Not the least its reliance on some kind of argument from analogy seems problematic (for an extensive criticism of the argument from analogy, cf. Avramides 2001). Ultimately, one needs to realize that there are other options available than the choice between theory-theory and simulation theory.
3. Gopnik and Wellman have compared the transition that occurs between the three-year old and the four-year old child's understanding of mind to the transition between Copernicus' *De Revolutionibus* and Kepler's discovery of elliptical orbits (Gopnik & Wellman 1995:242).
4. Occasionally, some of the theory-theorists have been cautious enough to admit that this parallelism might not hold true for all kinds of mental states, but there is no general agreement about what should count as the relevant exceptions.
5. For an informative overview of these tests, cf. Baron-Cohen (2000).
6. This is also granted by modularity nativism. Although this version of the theory-theory argues that the theory is innate, it still concedes that the theory needs a certain amount of experience as a trigger.
7. Certain core features in infantile autism have frequently been interpreted as a result of a mind-blindness, i.e., they have been explained by reference to a damaged or destroyed theory of mind mechanism. But if autists lack a theory of mind, and if a theory of mind is required for self-awareness, then autists should be "as blind to their own mental states as they are to the mental states of others" (Carruthers 1996b:262; cf. Frith & Happé 1999:1, 7). Thus, we find Baron-Cohen arguing that autistic subjects are "unaware of their own mental states" (Baron-Cohen 1989:595), and Frith and Happé proposing that persons with autism can only judge their own mental states by their actions (Frith & Happé 1999:11), i.e., denying that autists have a direct, immediate, or non-inferential access to their own mind.
8. Although Carruthers is, in general, unequivocal about denying conscious experiences to young infants (cf. Carruthers 1996:221, 2000:202–203), he occasionally leaves a door open for a different conclusion. As he writes at one point, it might be that infants are capable of discriminating between their experiences (and hence capable of enjoying conscious experiences), even while still being incapable of conceptualizing them (Carruthers 1996:222).

9. The problem is exacerbated by the fact that although “self-awareness” is an ambiguous term, the theory-theorists rarely *define* what precisely they mean when they speak of self-awareness (cf. Zahavi & Parnas 2003).
10. It could be objected that although this might hold true for humans, it does not necessarily hold true for all intelligent life. Would it for instance be nonsensical to imagine intersubjectivity between brains-in-vats or between disembodied angels? Is the very idea of telepathy incoherent? (Thanks to Galen Strawson for this objection.) A substantial reply would lead too far. But let me, on the one hand, simply confess that I am not all that convinced that it is legitimate to draw substantial philosophical conclusions from the fact that certain scenarios are imaginable. Is our imagination always trustworthy, does it always attest to metaphysical possibility, or might imaginability not occasionally reflect nothing but our own ignorance (for a more extensive discussion cf. Wilkes 1988; Parnas & Zahavi 2000)? On the other hand, I would insist that if something like intersubjectivity is possible between brains-in-vats or angels, then it is a kind of intersubjectivity that is utterly different from the one we are familiar with.
11. For more recent defenses of this position, cf. Flanagan (1992); Zahavi (1999, 2002, 2003); Kriegel (2003).
12. Let me forestall a possible objection, namely that this definition of self-awareness is too broad and that it simply includes too much. That is, since it doesn’t match our everyday or folk-psychological notion of self-awareness (that tends to link the notion with our ability to recognize or identify ourselves in a thematic way), the present use of the term is inappropriate. I don’t think this objection carries a lot of weight. From a conceptual point of view, there are no intrinsic problems whatsoever in using the term “self-awareness” to designate a situation where consciousness is aware of itself, or given to itself. Secondly, it is a simple fact that many of the classical philosophical theories of self-awareness as well as the more recent contributions by such thinkers as Brentano, Husserl, Sartre, Henry, Henrich, Frank, etc. have precisely been discussions of this broad notion. For a more extensive discussion, cf. Zahavi (1999).
13. For a more extensive overview of different phenomenological investigations of the body, cf. Zaner (1964); Leder (1990); Waldenfels (2000).
14. This study has been funded by the Danish National Research Foundation.

References

- Avramides, A. (2001). *Other minds*. London: Routledge.
- Baron-Cohen, S. (1989). Are autistic children ‘behaviorists’? An examination of their mental-physical and appearance-reality distinctions. *Journal of Autism and Developmental Disorders*, 19, 579–600.
- Baron-Cohen, S. (1995). *Mindblindness. An essay on autism and theory of mind*. Cambridge, MA: MIT Press.

- Baron-Cohen, S. (2000). The cognitive neuroscience of autism: Evolutionary approaches. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (second edition) (pp. 1249–1257). Cambridge, MA: MIT Press.
- Bermúdez, J. L. (1998). *The paradox of self-consciousness*. Cambridge, MA: MIT Press.
- Blackburn, S. (1995). Theory, observation and drama. In M. Davies & T. Stone (Eds.), *Folk psychology: The theory of mind debate* (pp. 274–290). Oxford: Blackwell.
- Butterworth, G. (2000). An ecological perspective on the self and its development. In D. Zahavi (Ed.), *Exploring the self. Philosophical and psychopathological perspectives on self-experience* (pp. 19–38). Amsterdam: John Benjamins.
- Carruthers, P. (1996a). Simulation and self-knowledge: A defence of theory-theory. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 22–38). Cambridge: Cambridge University Press.
- Carruthers, P. (1996b). Autism as mind-blindness: An elaboration and partial defence. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 257–273). Cambridge: Cambridge University Press.
- Carruthers, P. (1996c). *Language, thought and consciousness*. Cambridge: Cambridge University Press.
- Carruthers, P. (1998). Natural theories of consciousness. *European Journal of Philosophy*, 6, 203–222.
- Carruthers, P. (2000). *Phenomenal consciousness. A naturalistic theory*. Cambridge: Cambridge University Press.
- Carruthers, P. & P. K. Smith (1996). Introduction. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 1–8). Cambridge: Cambridge University Press.
- Davidson, D. (2001). *Subjective, intersubjective, objective*. Oxford: Oxford University Press.
- Fivaz, E. et al. (2004). Threesome intersubjectivity in infancy. In D. Zahavi, T. Grünbaum, & J. Parnas (Eds.), *The structure and development of self-consciousness: Interdisciplinary perspectives*. Amsterdam: John Benjamins.
- Flanagan, O. (1992). *Consciousness reconsidered*. Cambridge, MA: MIT Press.
- Frith, U. & F. Happé (1999). Theory of mind and self-consciousness: What is it like to be autistic? *Mind & Language*, 14, 1–22.
- Gallagher, S. (2003). Bodily self-awareness and object perception. *Theoria et Historia Scientiarum*, 7, 53–68.
- Gibson, J. J. (1979/1986). *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, 16, 1–14.
- Gopnik, A. (1996). Theories and modules: Creation myths, developmental realities, and Neurath's boat. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 169–183). Cambridge: Cambridge University Press.
- Gopnik, A. & H. M. Wellman (1995). Why the child's theory of mind really is a theory. In M. Davies & T. Stone (Eds.), *Folk psychology: The theory of mind debate* (pp. 232–258). Oxford: Blackwell.
- Gordon, R. M. (1996). 'Radical' simulationism. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 11–21). Cambridge: Cambridge University Press.

- Heal, J. (1996). Simulation, theory, and content. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 75–89). Cambridge: Cambridge University Press.
- Henry, M. (1965). *Philosophie et phénoménologie du corps*. Paris: PUF.
- Hobson, R. P. (1991). Against the theory of 'Theory of Mind'. *British Journal of Developmental Psychology*, 9, 33–51.
- Hobson, R. P. (1993). *Autism and the development of mind*. Hove: Psychology Press.
- Hobson, R. P. (2002). *The cradle of thought*. London: Macmillan.
- Husserl, E. (1973). *Zur Phänomenologie der Intersubjektivität II*. Husserliana XIV. Den Haag: Martinus Nijhoff.
- Kriegel, U. (2003). Consciousness as intransitive self-consciousness: Two views and an argument. *Canadian Journal of Philosophy*, 33, 103–132.
- Leder, D. (1990). *The absent body*. Chicago: University of Chicago Press.
- Legerstee, M. (1999). Mental and bodily awareness in infancy. In S. Gallagher & J. Shear (Eds.), *Models of the self* (pp. 213–230). Exeter: Imprint Academic.
- Leslie, A. M. (1987). Children's understanding of the mental world. In R. L. Gregory (Ed.), *The Oxford companion to the mind* (pp. 139–142). Oxford: Oxford University Press.
- Lewis, M. & J. Brooks-Gunn (1979). *Social cognition and the acquisition of self*. New York: Plenum Press.
- Mahler, M. S., F. Pine, & A. Bergman (1975). *The psychological birth of the human infant*. New York: Basic Books.
- McCulloch, G. (2003). *The life of the mind: An essay on phenomenological externalism*. London: Routledge.
- Merleau-Ponty, M. (1945). *Phénoménologie de la perception*. Paris: Éditions Gallimard.
- Merleau-Ponty, M. (1960). *Signes*. Paris: Éditions Gallimard.
- Merleau-Ponty, M. (1964). *Le visible et l'invisible*. Paris: Tel Gallimard.
- Neisser, U. (1988). Five kinds of self-knowledge. *Philosophical Psychology*, 1, 35–59.
- Neisser, U. (1993). The self perceived. In U. Neisser (Ed.), *The perceived self: Ecological and interpersonal sources of self-knowledge* (pp. 3–21). New York: Cambridge University Press.
- Nichols, S. & S. Stich (2002). Reading one's own mind: A cognitive theory of self-awareness. <http://ruccs.rutgers.edu/ArchiveFolder/Research%20Group/Publications/Room/room.html>
- Parnas, J. & D. Zahavi (2000). The link: Philosophy-psychopathology-phenomenology. In D. Zahavi (Ed.), *Exploring the self* (pp. 1–16). Advances in Consciousness Research. Amsterdam-Philadelphia: John Benjamins Publishing Company.
- Piaget, J. & B. Inhelder (1969). *The Psychology of the Child*. New York: Basic Books.
- Premack, D. & G. Woodruff (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 4, 515–526.
- Rochat, P. (2001). *The infant's world*. Cambridge, MA: Harvard University Press.
- Sartre, J.-P. (1943/1976). *L'être et le néant*. Paris: Gallimard.
- Scheler, M. (1973). *Wesen und Formen der Sympathie*. Bern/München: Francke Verlag.
- Shoemaker, S. (1968). Self-reference and self-awareness. *The Journal of Philosophy*, LXV, 556–579.
- Shoemaker, S. & R. Swinburne (1984). *Personal identity*. Oxford: Blackwell.

- Spitz, R. A. (1983). *Dialogues from infancy. Selected papers*. R. N. Emde (Ed.). New York: International Universities Press.
- Stern, D. N. (1983). The early development of schemas of self, other and 'self with other'. In J. D. Lichtenberg & S. Kaplan (Eds.), *Reflections on self-psychology* (pp. 49–84). Hillsdale: Analytical Press.
- Stern, D. N. (1985). *The interpersonal world of the infant*. New York: Basic Books.
- Strawson, P. F. (1959). *Individuals*. London: Methuen.
- Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), *Before speech. The beginning of interpersonal communication* (pp. 321–347). Cambridge: Cambridge University Press.
- Waldenfels, B. (2000). *Das leibliche Selbst. Vorlesungen zur Phänomenologie des Leibes*. Frankfurt am Main: Suhrkamp.
- Wilkes, K. V. (1988). *Real people. Personal identity without thought experiments*. Oxford: Clarendon Press.
- Wittgenstein, L. (1980). *Remarks on the philosophy of psychology II*. G. H. von Wright & H. Nyman (Eds.). Oxford: Blackwell.
- Zahavi, D. (1999). *Self-awareness and Alterity: A phenomenological investigation*. Evanston: Northwestern University Press.
- Zahavi, D. (2002). First-person thoughts and embodied self-awareness. Some reflections on the relation between recent analytical philosophy and phenomenology. *Phenomenology and the Cognitive Sciences*, 1, 7–26.
- Zahavi, D. (2003). Phenomenology of self. In T. Kircher & A. David (Eds.), *The self in neuroscience and psychiatry* (pp. 56–75). Cambridge: Cambridge University Press.
- Zahavi, D. & J. Parnas (2003). Conceptual problems in infantile autism research: Why cognitive science needs phenomenology. *Journal of Consciousness Studies*, 10, 155–176.
- Zaner, R. M. (1964). *The problem of embodiment. Some contributions to a phenomenology of the body*. The Hague: Martinus Nijhoff.

CHAPTER 4

From self-recognition to self-consciousness

Marc Jeannerod

Institut des Sciences Cognitives, France

1. Introduction

The first query is about ourselves: “What makes us self-conscious?”; or “What makes us such that we can consciously refer to ourselves as that particular self, different from other selves?” There are several ways to answer this question, according to the level at which one considers the idea of a self. One of these levels is that of the narrative self. As narrators, we obviously know who we are, where we are, what we are presently doing, and what we were doing before. Unless we become demented, we have a strong feeling of continuity in our conscious experience. We rely on declarative memory systems, from which souvenirs (albeit distorted) can be retrieved, and can be used as material for verbalization or imagination. Another level is that of the embodied self. We recognize ourselves as the owners of bodies and authors of actions. At variance with the narrative self, the type of self-consciousness that is linked to the experience of the embodied self is discontinuous: it operates on a moment-to-moment basis, as it is bound to particular bodily events, like actions, for example. Instead of explicitly answering questions like: “Who am I?”, something that the narrative self needs to know permanently, the embodied self will answer questions like, “Is this mine?” or “Did I do this?”, which are questions to which we rarely care to give explicit responses. In other words, the embodied self mostly carries an implicit mode of self-consciousness, whereby self-consciousness is around, but becomes manifest only when required by the situation. The related information has a short life span and usually does not survive the bodily event for very long.

This chapter will investigate one essential aspect of self-consciousness at the embodied level, self-recognition, i.e., the ability to recognize oneself as the owner of a body and the agent of actions. This ability requires specific mecha-

nisms, which have been elucidated only recently. First, the matching of visual, tactile, and proprioceptive signals originating from the same body parts contributes to an intermodal sensory image of the body. Second, the monitoring of one's intentions contributes to a sense of the self as an agent. The congruence between these two sets of signals is a strong index for determining the experiences of ownership and agency, which are the main constituents of the experience of being an independent self.

2. The role of sensory cues in self-recognition

Sensory cues contribute to the building of the body image, i.e., a representation of the whole body to which body parts can be referred (e.g., Gallagher 2000). As stressed by many experimental results, vision has a dominant role over other senses in this process: we feel our hand where we see it, not the converse. Optical distortion of the visually perceived position of a limb with respect to its felt position (e.g., by wearing laterally displacing prisms) produces no alteration of the sense of ownership: the position sense is actually recalibrated to conform to the visual information (Harris 1965). This predominance of vision was confirmed in experiments using a rubber hand. Botvinick and Cohen (1998) positioned a realistic rubber arm in front of subjects, while their real arm, hidden by a screen, was placed aside: tactile stimulation was applied simultaneously to the real and the rubber arms. After some time, the subjects experienced an illusion in which they felt the touch at the locus of the rubber arm (that they could see), not of their real (hidden) arm. In other words, the tactile stimulus was felt at the place where it was seen, at the expense of a distortion of the felt position of the real arm. In addition, subjects spontaneously reported experiencing a clear sense of ownership for the rubber arm. According to other authors who replicated this experiment, the illusion of displacement of the tactile stimulus and the illusion of ownership disappear if the rubber arm is not properly aligned with the subject's body (Farné et al. 2000). Indeed, simply looking at a moving limb superimposed to one's own limb creates a strong impression of having willed this movement and of being its author. Observations have been reported in amputated people who experience having a phantom limb. When their valid limb is visually transposed (by way of mirrors) to the amputated side, and when they produce movements with that limb, they experience a strong feeling of voluntary movement of the phantom limb. The same happens if the visually transposed limb is that of an experimenter (Ramachandran & Rogers-Ramachandran 1996). As argued by Wegner (2002),

these observations reflect a tendency to perceive oneself as causal. Even if what we experience reflects a simple (visual) appearance of mental causation, when we experience it, we tend to attribute the cause to ourselves.

Body ownership, however, is only part of the problem of self-recognition. The self is most of the time an acting self. Body parts are moving with respect to one another and with respect to external objects as the result of intentional actions. It is common experience that our actions are readily self-attributed as a consequence of a normally perfect correlation between their expected effects and the flow of resulting (visual and proprioceptive) stimulation. This matching process provides the agent of an action with the sense that he is causing that action (the sense of agency). As it is illustrated by a famous movie scene, an efficient means for determining ourselves as the owners of our bodies and body parts, as we see them, is to make them move: if the image I see in a mirror in front of me moves when I move and the two movements are congruent, then the image I see must belong to me. Babies at a very early age seem to use this congruence criterion for attributing to themselves their body parts. Developmental studies provide evidence that self-recognition appears early in life. Infants at five months of age are able to discriminate their own leg movements displayed in a mirror from those of another infant, presumably by making use of a perceived contingency between their own behavior and its effects (Bahrnick & Watson 1985). As they grow older, the infants' behavior will increasingly testify to their development of a conscious self-representation. Infants of 15–20 months of age, for example, will typically resolve the task of wiping a red spot stuck on their faces, when they see themselves in a mirror (see Bahrnick 1995, for review). Below, we will examine experiments where this notion of congruence was measured and manipulated.

Again, however, situations may arise where this attribution becomes less than obvious. In social interactions, several people may participate in the same action and interact rapidly on the same object. Playing ball games is one example. Another one is that of two surgeons operating jointly in the same surgical theater and seeing their respective hands through a magnifying lens. In both cases, there are several moving hands visible in the scene. These hands may not appear to be directly connected to the corresponding body, as in the surgical example. Yet, these movements and the corresponding hands are correctly attributed to their authors. What is meant by these examples is that attributing to oneself both the ownership of a body part and the authorship of a movement must be based on specific mechanisms, which, in everyday life, are sufficiently accurate to allow unambiguous self recognition. Both the sense of ownership and the sense of authorship concur to self-identification: it is as essential to

recognize oneself as the owner of one's body as it is to recognize oneself as the agent of one's actions.

3. The Nielsen paradigm for studying the recognition of self-generated actions

The rest of the chapter will be devoted to describing the cues used for recognizing one's own actions. A set of pioneer experiments will be reported first, as they have set the stage for more recent investigations of the sense of agency. These experiments were undertaken in the 1960s by Torsten Nielsen working at the Psychological laboratory of the University of Copenhagen.¹ One of the phenomena Nielsen thought of great importance for self-awareness was the volitional experience, that is, the experience of volitional or intentional control of perceived events, as opposed to the experience of no control of those events. Nielsen's approach to examining this issue was to create situations in which the experiences of intentional control vs lack of control could be experimentally manipulated. To this aim, he created several varieties of a *substitution* paradigm, where the subjects received false feedback from their own actions: in fact, what the subjects perceived was the effect of the actions of another person, which was substituted for their own.

In one of Nielsen's experiments (1963), the experimental subject was facing a box placed on a tabletop. He placed his hand holding a pencil on the table below the box, and looked at it through the box. The box was equipped with a mirror, which could be displaced by the experimenter between trials, without the subject's awareness, so that the subject either saw his own hand or the hand of another person (the alien hand) through the mirror. The mirror was placed in such a way that the subject's experience was that he was looking directly at his own hand, while in reality he was presented with the alien hand, lying at the same location on the table as his own. Finally, to ensure that the subject had no cues to identify the hand he saw, the two hands were made undistinguishable by wearing identical gloves. During the experiment itself, the subject was requested to draw a straight line in the sagittal direction on a piece of paper (Figure 1). In those trials where the subject saw the alien hand, the alien hand was also doing the same task at about the same rate. In some trials, however, the alien hand carried out a movement that diverged from that the subject was carrying out at the same time. The latter condition generated a conflict between what the subject saw and what he kinesthetically felt from what he was doing. In order to solve this conflict, the subjects tended to deviate the trajectory of

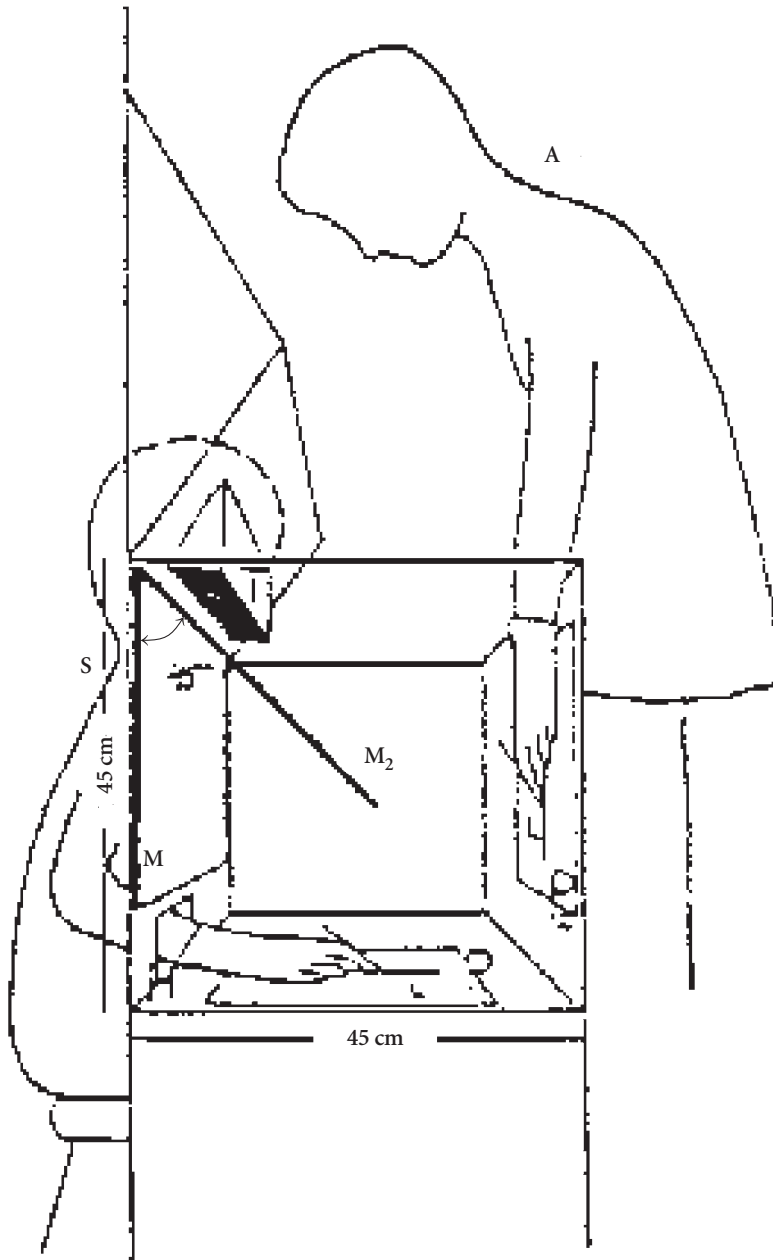


Figure 1.

their own (unseen) hands in the direction opposite to that of the alien hand, so as to fulfill the instruction they had received to draw a straight line.

According to Nielsen, all the subjects in the conflict trials experienced that they saw their own hands moving involuntarily in the wrong direction. They remained unaware of having themselves performed a movement departing from the instruction, by erring in the direction opposite to that of the alien hand, to compensate for the conflict produced by the movement they saw: when shown their own deviant performance, they tried to explain it by factors independent from their volition, such as fatigue or inattention. In Nielsen's terms, some subjects reported impressions of loss of voluntary control, as if driving one's car on an icy road. Thus, this experiment revealed that subjects were poor at recognizing their own hand movements and tended to misattribute to themselves movements that were not theirs. The cues arising from the visual perception of the hand of the other person dominated the kinesthetic cues arising from the subject himself in determining self-awareness.

Two years later, Nielsen together with his colleagues, published the results of a second experiment based on the same substitution paradigm, but using volitional performance in a different domain, that of vocal utterance. In this experiment (Nielsen et al. 1965), subjects were equipped with headphones. A note at around 400 Hz was displayed to them by a sine-wave generator and they were instructed to match this note by singing. After 3 seconds, the generator was turned off and the subjects heard their own voices. They were instructed to hold the same note for several seconds. On some trials, the subject's voice was replaced by the voice of an assistant initially singing at the same frequency but then continuously falling by 10 to 30 Hz. This situation generated from the subjects a 'compensatory' increase in the frequency of the their own singing. The rise in frequency of the subject's voice was about equal or even greater than the fall in frequency of the assistant's voice. When asked to describe their experiences during those trials, nearly all subjects commented that they were hearing their own voices with an unexplained falling pitch. Furthermore, when listening to the recordings of their own performance, they were surprised to hear the rising pitch of their voices and reported that they were not aware of doing this during the trials. The same compensatory behavior was observed in the reverse situation, i.e., when the subjects heard an assistant's voice with a continuously increasing pitch: they decreased their own singing frequency and, even after hearing their performance, they did not give up the idea that, during the trial, they had heard their own voices with a rising pitch. As a rule, the subjects experienced their own performance as involuntary and mentioned that they tried to correct the changing pitch of the voice, but could not succeed.

Nielsen et al.'s conclusion was that the subjects had no distinct kinesthetic awareness of their own vocal movements and that "the auditory voice perception dominated the kinesthetic voice awareness" (1965:206). Concerning the experience of loss of voluntary control reported by the subjects, both in the experiment with hand substitution and in the voice substitution experiment, Nielsen stated: "Usually, people are not explicitly aware of volitional aspects of their own performance until they meet difficulties in carrying out an intentional act. The present experiments created difficulties in such a way that they suddenly became conscious of the volitional aspects of singing a note" (1965:208).

Finally, Nielsen performed another experiment, which he reported in a treatise in Danish with a long summary in English (Nielsen 1978). This experiment basically transferred the previous substitution paradigm to the movements of the whole body. The subject standing on a platform looked at a reflection of his body in a mirror placed below his feet and was instructed to sway his body straight ahead. On some trials, the mirror was unknowingly removed and the subject was now looking down at a dummy (an 'alien body'): when the subject moved straight ahead, the alien body was made to move in a different direction, e.g., to the right. Subjects consistently reported that they experienced that their own bodies were pulled to the right. As stated by Nielsen: "When the dummy swayed toward the right, [subjects] were quite certain that they were looking at themselves, and the visual impression actually shaped their postural body consciousness to such a degree that they also felt that they were being forced over to the right, against their will" (1978:259). Indeed, some subjects moved to the left in an attempt to rectify the rightward movement.

Nielsen's experiments were embedded in a conceptual framework aiming at a formal description of how human beings can act and reflect on their acts. His main idea was to present an alternative to psychological theories, which are based on a deterministic and objectified view of human action. Human beings, he thought, rather than the psychologist, must provide themselves answers to the questions of how they understand their own actions. By placing subjects in situations where everyday life activities become suddenly incomprehensible and unfeasible, Nielsen thought that they would feel compelled to ask themselves 'action-clarifying questions' and that these spontaneous questions "would generate a useful and inspiring material for [his] efforts to constitute some basic concepts concerning action..." (1978:258).

Nielsen's endeavor generated little insight in psychological circles until recently, when his substitution paradigm was systematically reinvestigated (see below). In 1969, however, Sullivan replicated the hand experiment of Nielsen

(1963) with the idea that the conflict between sensory cues from different modalities produced by the substitution of the subject's hand by an alien hand could be a model for explaining the denial of the effects of stroke in some patients. Patients experiencing an impairment in integration of visual and somatosensory cues following a stroke are likely to believe that their senses still provide them with veridical information about their body, as this was the case before the stroke. During the initial post-stroke period, these patients presumably rely on visual cues to test out their perceptions (a likely hypothesis following Nielsen's finding of visual dominance over somatosensory cues). Because the patients can see, e.g., their arms, they will assume that they are not impaired and will finally come to deny their illness. Sullivan's experiment with normal subjects confirmed the dominance of visual cues (and correlatively the neglect of kinesthetic cues) in shaping the subjects' experiences. Subjects believed that they had drawn straight lines and expressed their surprise when they saw their deviated lines. They made remarks about their experiences that were similar to those made by some stroke patients during the initial period. According to Weinstein and Kahn's (1959) classification of responses of patients with anosognosia, some of these remarks were typical somatagnosic² responses or even denial responses (Sullivan 1969).

4. Recent experiments using the Nielsen substitution paradigm

Nielsen left several issues unanswered. One of them was the limit to which the substitution operates in fooling the subject: until what point will a subject accept the alien hand as his own? In order to explore specifically this point, the Nielsen hand experiment was replicated using several slightly different versions. In an experiment by Fournieret and Jeannerod (1998), subjects were instructed to move a stylus in the direction of a visual target with their unseen hand: only the trajectory of the stylus was visible as a line on a computer screen, seen in a mirror superimposed to the subject's hand. In some trials, an angular bias of a variable amplitude (e.g., 2°, 5°, or 10° to the right or to the left) was introduced electronically, such that the visible trajectory no longer corresponded to that of the hand. In order to reach the target, the hand-held stylus had to be deviated in a direction opposite to the bias. In other words, although the line on the computer screen appeared to be directed to the target location, the hand movement was directed in a different direction. At the end of each trial, subjects were asked in which direction they thought their hands had moved.

This experiment revealed several important points. First, subjects accurately corrected for the bias by tracing a line that deviated by the same angle as the bias, but in the opposite direction (what Nielsen had called a compensatory behavior). Second, subjects tended to ignore the veridical trajectory of their hands in making a conscious judgment about the direction of their hand movements: instead, like Nielsen's subjects, they based their reports on visual cues and tended to adhere to the direction seen on the screen, thus ignoring non-visual (e.g., proprioceptive) cues. The results therefore suggest that the visuomotor system is able to appropriately use information for producing accurate corrections, but that this information cannot be accessed consciously. However, when the bias is progressively increased from trial to trial above the limit of 10° that was used in the Fournier and Jeannerod experiment, the subjects suddenly change strategy (at an average value of bias of 14° , Slachevsky et al. 2001) and begin to use conscious monitoring of their hand movements to correct for the bias and to reach the target. In other words, the discrepancy between the seen trajectory and the felt trajectory becomes too large to be automatically corrected, and the discordance observed by the subjects has to be compensated by conscious deviations of the hand movement in the appropriate direction.

Although this series of experiments dealt with recognition of action and with the degree of awareness that a subject can gain from his own movements, it said little about the cues that can be used for conscious determination of agency. This was the objective of another set of experiments initiated by Daprati et al. (1997), which explored the factors of self attribution of a moving hand. A situation was created in which the subjects were shown movements of a hand of an uncertain origin, that is, a hand that could equally likely belong to them or to someone else. The subject's hand and the experimenter's hand, wearing identical gloves, were filmed with two different cameras. By changing the position of a switch, one or the other hand could be briefly (5 seconds) displayed on a video screen. The subject saw the screen through a mirror placed above her own right hand. Thus, looking at the mirror, the subject got the impression that she was watching her own hand.

The task for the subject was to perform a requested movement with her right hand, and to monitor its execution by looking at the image in the mirror. At the beginning of each experimental trial, a blank screen was presented. An instruction to perform a movement was given and the subject and the experimenter had to execute the requested movement at an acoustic signal. Once the movement was performed and the screen had returned blank, the subject was asked whether the hand that she just saw was hers or not. One of three possible

images of the hand could be presented to the subjects in each trial: 1. the subject's own hand (condition: Subject); 2. the experimenter's hand performing a different movement (condition: Experimenter Different); 3. the experimenter's hand performing the same movement (condition: Experimenter Same).

Subjects were able to unambiguously determine whether the moving hand seen on the screen was theirs or not, in two conditions. First, when they saw their own hands (trials from the condition Subject), they correctly attributed the movements to themselves. Second, when they saw the experimenter's hand performing a movement that departed from the instruction they had received (condition Experimenter Different), they denied seeing their own hands. By contrast, their performance degraded in the condition Experimenter Same, that is, in trials where they saw the experimenter's hand performing the same movement as required by the instruction: in this condition, they misjudged the hand as theirs in about 30% of cases. Subjects' judgment had to rely on slight differences in timing and kinematics between their intended movements and those they perceived on the screen. This result therefore indicates that the threshold for action recognition must be relatively high and that small differences tend to be neglected. This is consistent with the above observation of Fournieret and Jeannerod (1998), where a small discordance between the movement and its visual consequence was ignored.

5. Self recognition in the social context

One of the key issues of the experiments performed by Daprati and her colleagues was that normal subjects who misrecognized the hand shown to them tended to overattribute that hand to themselves. One possible explanation for this effect could be that subjects saw only one hand and because they knew they had moved their fingers during the presentation, the hand was automatically attributed to the author of the movements.

To avoid this possible confound, a different situation was used, which combined uncertainty about ownership of the subject's hand and uncertainty about authorship of the movements performed with that hand. This situation, designed by C. Farrer, involved simultaneous presentation of two hands, one of which was the subject's hand, the other being an alien hand. This situation is more realistic than the one used in previous experiments, since it involves 'social' interaction between two people, in which problems of self vs other recognition are more likely to arise. The question in this situation was therefore not whether an observed action corresponded to the action one had performed,

but rather which of two observed actions was the one corresponding to the action performed by the self.

The subject and the experimenter sat at the opposite sides of a table. The subject was facing an LCD screen. Both the subject and the experimenter placed their right gloved hands below the screen. A mirror attached to the backside of the screen reflected the image of the two hands to a video camera connected to a computer. A program processed the digitized video image in real time (within 20 ms) and sent an image of the hands onto the LCD screen. The program allowed the rotation of the image displayed on the screen by -90° , 90° , and 180° . So, the subject could see his own hand at the bottom of the screen, where it would be in reality (0° Rotation), at the top of the screen (180° Rotation), at the left of the screen (90° Rotation), or at the right of the screen (-90° Rotation), while the experimenter's hand was always in the opposite direction. The rotations provided the possibility to study how the recognition of one's hand was influenced by the location where it appeared visually.

At the beginning of each trial, the subject was instructed to either extend his index finger or his thumb, or to make no movement. During the trials where the subject was instructed to make a movement, the experimenter would either make the same or the alternative movement. Once the movements were performed, the screen returned dark within about 1s. Then a pointer was placed at the position where one of the two hands had been. Subjects had to determine whether the hand indicated by the pointer was theirs or that of the experimenter.

This experiment (van den Bos & Jeannerod 2002; see also Knoblich 2002), first, allowed the experimenters to study the role of the apparent positions of the hands concerning self-recognition. When the two hands visually appeared at the loci corresponding to their real positions, subjects showed relatively little difficulty in recognizing their own hands. However, when the apparent locations of the hands were interchanged with respect to reality, they made attribution errors. This confirms that the contingency between visual and proprioceptive signals plays a role in self-recognition. However, the most critical factor for correct attribution was the presence of finger movements. When finger movements were present and these movements were clearly attributable to the self (i.e., they differed from those of the experimenter), no attribution errors occurred. This result replicates Daprati et al.'s (1997) finding that subjects correctly attributed the hand they saw when the finger movements were theirs or when the hand was that of the experimenter performing different movements. The surprising finding in the present experiment is that accurate self-recognition was possible for all orientations of the display, including the

180° rotation. In other words, when distinctive movements are available, subjects tend to recognize actions, not just hands. By contrast, when the two hands performed the same movements or no movements, the error rate increased as a function of the degree of rotation. Finally, another interesting aspect of the results was the direction of errors (e.g., misattribution to the self or to the other person): when movements were not discriminative (e.g., when they were the same or absent), subjects misattributed the indicated hand more often to themselves than to the other. This result, together with those of the other experiments using the Nielsen paradigm reveal that, in conditions where signals for self-recognition are weak, subjects tend to overattribute the movements to themselves.

6. The nature of the mechanisms involved in action recognition

There are several ways of conceiving the mechanisms involved in action recognition. One hypothesis relies on the idea that executed actions generate signals that are centrally monitored and compared: action recognition arises as the outcome of this comparison (the central monitoring hypothesis). The other hypothesis relies on the idea that actions, whether or not they come to execution, are centrally simulated by the neural network, and that this simulation is the basis for action recognition (the simulation hypothesis). These two theories appear to be largely complementary.

6.1 The central monitoring hypothesis of action recognition

The first hypothesis to be considered, which holds that the comparison between efferent signals at the origin of an action and those that arise from its execution (the reafferent signals) provides cues about where the action originates, is deeply rooted in physiological thinking. Let us first rephrase its basic principles. The original idea, inherited from the cybernetic era and still operational nowadays, is that each time the motor centers generate an outflow signal for producing a movement, a copy of this command (the “efference copy”) is retained. The reafferent inflow signals generated by the movement (e.g., visual, proprioceptive) are compared with the copy (Sperry 1950; von Holst & Mittelstaedt 1950). An essential aspect of this mechanism is its predictive nature. It is based on “internal models”, where the states of the motor system are represented: the current state of the system, its desired state, and its predicted state. Within the internal model, predictors capture the causal rela-

tionships between actions and their consequences, whereas controllers provide the necessary motor commands. Prediction in a sensorimotor system is needed to evaluate sensory information resulting from a movement. The sensory consequences of a self-produced movement can be accurately predicted from the efference copy generated in parallel with the motor command: there should normally be little discrepancy between the predicted and the actual sensory feedback. In contrast, sensations arising from the outside cannot be predicted and will result in a higher degree of discrepancy. The same stimulus will therefore be perceived differently whether it arises from a self-produced movement or it is externally produced (see Wolpert et al. 1995; Blakemore et al. 1998).

This model of the control of action can be directly applied to the problem of self-recognition. Self-recognition can be based on the concordance between a desired (or intended) action and its sensory consequences, which can be used to assess attribution of action to the self. This hypothesis can be tested experimentally. Haggard et al. (2002), using a paradigm initiated by Libet et al. (1983), instructed subjects to make a simple voluntary movement (a key press) at a time of their choice. The action of pressing the key caused an auditory signal to appear after a fixed delay of 250 ms. In separate sessions, the subjects were asked either to report the position of a clock hand at the time they thought they had pressed the key, or at the time when they heard the auditory signal. Haggard et al. found that the time interval between the two estimated events was shorter than expected, i.e., 250 ms. Subjects tended to perceive their key press occurring later, and the auditory signal occurring earlier, than it was actually the case. This shrinkage of perceived time between the two events did not happen in a control situation where the finger movement was not voluntary, but was produced by a magnetically induced stimulation of motor cortex. The authors conclude that intentional action binds together the conscious representation of the action and its sensory consequences. This binding effect would thus account for normal subjects' self attribution for their own actions.

The action monitoring model capitalizes on peripheral signals produced by the subject's motor activity. It postulates that these signals are used for comparison with the internal model of the action. There are many situations, however, where an action representation is formed, but no movement is executed. In such situations, no output signals to the muscles, no reafferent (e.g., visual) signals from the outside world, no proprioceptive signals (and therefore, no possibility for comparing execution with a desired output) exist. Yet, the attribution of the representation is clearly made to the self. The self/other distinction of the origin of an action may seem relatively simple when movements

are overtly executed. In this condition, there is a possibility for monitoring sensory signals arising from the moving limbs and from the effects on the external world of the movements produced either by the self or the other, and comparing them with the signals resulting from the action generation mechanism. This distinction is made difficult in a number of situations where the action generation mechanism is activated, but where the action remains covert. In the section below, we will consider two such situations. One is that of imagined actions, which have been extensively studied under the heading of “motor imagery”; another one is that of observation of actions performed by someone else, a situation that has been introduced more recently within the realm of covert action. Although the two situations are clearly dissimilar, they both rely on neural simulation of the imagined or the observed action.

The existence of covert action, which is an essential constituent of motor cognition, raises the problem for the self of disentangling from one another different modalities of motor representations that may be present more or less simultaneously within the same brain. Action recognition thus cannot be treated separately from the more general process extending to recognition of action-related mental states like intentions or desires.

6.2 The simulation theory: From motor imagery to action attribution

The notion that a movement needs to be executed for the self-recognition mechanisms to operate is a problem in itself. Action often remains within the mind without becoming apparent. The existence of overt behavior is not a prerequisite for self-attribution. It has been often argued that thinking, which normal subjects unambiguously attribute to themselves, is an equivalent of a weak form of behavior that does not activate muscles and is therefore invisible from the outside (see Feinberg 1978; Hesslow 2002; Wegner 2002). A paradigmatic situation where this occurs is motor imagery. In this section, we will describe recent research on motor imagery and look for another explanatory model of action recognition, attribution, and self-recognition based on neural simulation of actions.

Our hypothesis thus postulates that covert actions like motor images are in fact actions in their own right, except for the fact that they are not executed. Covert and overt stages represent a continuum, such that every overtly executed action implies the existence of a covert stage, whereas a covert action does not necessarily turn into an overt action. Most of the neural events that lead to an overt action already seem to be present in the covert stages of that action (e.g., Jeannerod 1994, 1999). Thus, a covert action includes everything

that is involved in an overt action, except for the muscular contractions and the joint rotations. Even though this contention is factually incorrect, as we know that the musculo-articular events associated with a real movement generate a flow of (reafferent) signals that are not present as such in a covert action, it captures the functioning of the representation. The theory therefore predicts a close similarity, in neural terms, of the state where an action is simulated and the state that immediately precedes execution of that action.

Early interest in the field of mental imagery, more than a century ago, was evident in the context of hypnosis. To explain hypnotic phenomena, authors of the time (e.g., Binet 1886) claimed that mental images in general resulted from excitation of the same cerebral centers as the corresponding actual sensation. In the domain of motor images, it was remarked that the state of the motor centers influence the possibility of generating a motor image. For example, it was shown to be impossible for a subject to generate the image of pronouncing the letter /b/ if he kept his mouth wide open: this was because, supposedly, the motor system cannot be engaged in two contradictory actions at the same time. More recently, imagined actions or motor images have become a major tool for the study of representational aspects of action. Among the most impressive behavioral findings revealed by motor imagery studies is the fact that motor images retain the same temporal characteristics as the corresponding real action when it comes to execution. For example, it takes the same time to walk mentally to a prespecified target as it takes to actually walk to the same place (Decety et al. 1989). Similarly, temporal regularities that are observed in executed actions are retained in their covert counterparts (Decety & Jeannerod 1996; Sirigu et al. 1996). Along the same line, other situations have been described where the subject is requested to make a perceptually-based “motor” decision. In this situation, no conscious image is formed. Consider for example, a subject who is asked to make an estimate about the feasibility of an action, e.g., to determine the feasibility of grasping an object placed at different orientations: the time to give the response is a function of the object’s orientation, suggesting that the arm has to be mentally moved to an appropriate position before the response can be given. Indeed, the time to make this estimate is closely similar to the time it takes to actually reach and grasp an object placed at the same orientation (e.g., Frak et al. 2001; see also Parsons 1994; Johnson 2000).

This indication of a similar temporal structure for executed and non-executed actions by a biological system is reinforced by a similarity at the level of physiological indicators. Examining autonomic activity in subjects imagining an action at different effort rates reveals changes in heart rate and

respiration frequency proportional to the imagined effort, in the absence of any metabolic need. These results (Decety et al. 1993; see review in Jeannerod 1995) reveal the existence of a central patterning of vegetative commands during covert actions that seem to parallel the preparation of muscular commands. Autonomic changes occurring during motor imagery are closely related to those observed during central preparation for an effortful action. These mechanisms appear to anticipate forthcoming metabolic needs, with the function of shortening the intrinsic delay required for heart and respiration to adapt to effort.

Research on motor imagery has thus reinforced the notion of action simulation, according to which, covert actions are mentally (and neurally) simulated by the potential agent. Hence, supporting the above hypothesis that a covert action should present the same features as a real action. The other form of covert action that was mentioned earlier in this chapter, namely observation of action performed by another person, should also be liable to the same explanation. Indeed, it has often been claimed that we understand the minds of others by simulating them, an idea that was already present in the literature a century ago under the concept of empathy (Lipps 1903). Empathy expresses the possibility that we understand other people's behavior (e.g., their actions, their facial expressions, etc.), because we attempt to replicate and simulate their mental activity. In other words, the observed action would activate, in the observer's brain, the same mechanisms that would be activated if that action was intended or imagined by the observer (e.g., Gallese & Goldman 1998).

Recent experiments have confirmed that action simulation can be a robust cue for recognizing both one's own actions and those observed of other people. In a study by Knoblich and Flach (2002), subjects were shown videos of an action (throwing darts) that they had previously performed and videos of the same action performed by other subjects. The subjects' task was to predict the accuracy of the observed actions. Prediction was more accurate when subjects observed their own actions than when they observed another person's actions. This result suggests that the observation of self-generated actions is more informative, because the pattern of the observed action was produced by the same mechanism that simulates the action during the observation (see Knoblich & Prinz 2001).

The similarity in mechanisms underlying action generation and action observation accounts for the fact that action attribution remains a fragile process. Indeed, in everyday life there are ambiguous situations in which the cues for the sense of agency become degraded and that obviously require a subtle mechanism for signaling the origin of an action. As the experiments above

have shown, situations can be created where normal subjects fail to recognize their own actions and misattribute to themselves actions performed by another agent. This may also be the case of situations created by interactions between two or more individuals (e.g., joint attention, matched actions, or mutual imitation), or situations pertaining to the domain of human-machine interactions (e.g., telemanipulation, virtual reality systems, etc.).

6.3 Differences and similarities between the two theories

The simulation theory and the central monitoring hypothesis described in the previous two sections have a number of points in common. In one of its latest versions, the central monitoring hypothesis postulates that the internal model at the origin of an action can operate on a purely representational basis. According to Frith et al. (2000), the internal model can estimate the current state of the motor command of a movement to predict the next state, by simulating the movement dynamics. It can also predict the sensory feedback that would result from the movement if it were executed. The sensory error – the difference between actual and predicted sensory feedback – can be used to correct the state estimate.

Although both theories involve some degree of simulation, however, they clearly depart from each other concerning the nature of the represented actions. As was stressed above, recognition of action and self-recognition are ultimately processes involving the participation of several persons, a social process. The main point is that actions of others are represented to the same extent as one's own, and because the two types of representations may be present in the same brain, they must be disentangled from each other. Self-recognition mechanisms thus cannot operate from a solipsist point of view, they must consider the existence of other agents. In the recent model proposed by Frith (Frith et al. 2000), it is suggested that self recognition could be achieved through the interpretation of the signals related to an action: if these signals do not belong to the self, they must belong to someone else (a default attribution). In the new model that is proposed here, specific mechanisms for attribution to another agent are postulated.

Extending the simulation theory to understanding the representations underlying actions of other people requires that a continuity is established between the embodiment of representations of the observer and those of the agent being observed. A simple experiment with normal subjects sheds light on this issue. Consider a subject observing the action of an actor. The subject's respiration rate is recorded while she sits in front of a large screen, on

which she sees an actor performing an effortful action. The actor stands on a treadmill that is either motionless or moves at a constant velocity (2.5, 7 or 10 km/h), or progressively accelerates from 0 to 10 km/h over one minute. The main result of this experiment (Paccalin & Jeannerod 2000) is that the respiration rate of the observer increased during the observation of the actor walking or running at an increasing speed. Typically, the average increase during observation of the actor running at 10 km/h is about 25% above the resting level. Further, the increase in respiration frequency of the observer correlated with running velocity. Watching an action is thus different from watching a visual scene with moving objects. While watching an action, the observer is not only seeing visual motion, he is also internally (and non-consciously) simulating (or rehearsing) the action. The simulation of accelerating running should cause an increase of the breathing rate, because if the running movements were actually executed, they would require an anticipatory increase in metabolic needs. This finding thus substantiates the hypothesis that perceiving an action triggers a neural state, where the neural structures potentially involved in executing that action are facilitated (see details below). Experiencing and watching an action would thus be the two faces of the same phenomenon.

7. A neural hypothesis for self-recognition and its failures: The 'Who' system

In this concluding section, a framework for integrating self-recognition with the neural substrate and for accounting for its failures is presented. The conception of action recognition defended in this chapter (Georgieff & Jeannerod 1998, 1999) is based on the existence of neural networks subserving the various forms of representation of an action. Although these ensembles are clearly distinct from one form of representation to another (e.g., the representation of a self-generated action *vs* the representation of an action observed or predicted from another agent), they are likely to partly overlap (see Jeannerod 2001). When two agents socially interact with one another, this overlap creates *shared representations*, i.e., neural networks that are simultaneously activated in the brains of the two agents. In normal conditions, however, the existence of non-overlapping parts, as well as the existence of possible differences in intensity of activation between the activated zones allow each agent to discriminate between representations activated from within from those activated from outside, and to disentangle which belongs to him from that which belongs to the other. This process would thus be the basis for correctly attributing a repre-

sentation (or the corresponding action) to its proper agent or, in other words, for answering the question of “Who” is the author of an action. Each agent builds in his brain a representation of both his own intended actions, using internal cues like his own beliefs and desires, and the potential actions of the other agent with whom he interacts. These partly overlapping representations are used by each agent to build a set of predictions and estimates about the social consequences of the represented actions, if and when they would be executed. Indeed, when an action comes to execution, it is perceived by the other agent as a set of social signals that confirm (or not) his predictions and possibly modify his beliefs and desires.

This conception allows us to hypothesize about the nature of the dysfunction responsible for misattribution of actions in pathological conditions. Changes in the pattern of cortical connectivity could alter the shape of the networks corresponding to different representations, or the relative intensity of activation in the areas composing these networks. A specific configuration of symptoms appearing in schizophrenic patients has been the target of intensive search in recent years. These so-called “positive” symptoms include verbal hallucinations, insertion of thought, and delusion of influence (Schneider 1955), and should represent mechanisms that normally specify the boundaries between the self and other people. Patients with positive symptoms tend to overattribute to themselves actions performed by others or, conversely, to attribute their own actions or thoughts to the influence of others. When placed in experimental situations like those already described in the previous sections, these patients have been found to systematically misattribute their movements or those of the experimenter (Daprati et al. 1997; Franck et al. 2001; Blakemore et al. 2000).

One of the possible explanations for this failure to recognize one’s own actions and to correctly attribute actions to their agents has focused on the role of the prefrontal cortex. The prefrontal cortex is known to be hypoactive in many schizophrenic patients (Weinberger & Berman 1996). Its morphological aspect has also been shown to be modified on post-mortem examination (Goldman-Rakic & Selemon 1997). Because prefrontal areas are known to normally exert an inhibitory control on other areas involved in various aspects of motor and sensorimotor processing, an alteration of this control in schizophrenic patients might result in aberrant representations for actions. Indeed, neuroimaging studies have revealed that patients presenting verbal hallucinations (e.g., “voices” whereby inner speech is erroneously attributed to external agents) show abnormal activation of primary auditory areas in their left temporal lobe, as if they were processing an external auditory stimulus

(Dierks et al. 2000). Similarly, an increased activity in the right posterior parietal lobe has been observed in patients with delusion of influence, either at rest (Frank et al. 2002) or during an action recognition task (Spence et al. 1997). Such increased activation could likely be related to the loss of frontal inhibition in the corresponding areas. It therefore can be suggested that an alteration in the pattern of connectivity of these networks would change the degree of overlap between the representations of self-generated and observed actions in such a way that the representations would become undistinguishable from each other. The pattern of misattribution would be a direct consequence of this alteration.

Notes

1. I am indebted to J. B. Sørensen for introducing me to the laboratory where T. Nielsen was working and for allowing me to examine the hand apparatus used in the 1963 experiment.
2. The term *somatagnosia* used here by Sullivan can be translated into ‘failure to recognize part of one’s body’. Neurologists refer to this impairment as *autotopagnosia*.

References

- Bahrick, L. E. (1995). Intermodal origins of self-perception. In P. Rochat (Eds.), *The self in infancy. Theory and research* (pp. 349–373). Amsterdam: Elsevier.
- Bahrick, L. E. & J. S. Watson (1985). Detection of intermodal proprioceptive-visual contingency as a potential basis of self-perception in infancy. *Developmental Psychology*, 21, 963–973.
- Binet, A. (1886). *La psychologie du raisonnement. Recherches expérimentales par l’hypnotisme*. Paris: Alcan.
- Blakemore, S. J., D. Wolpert, & C. D. Frith (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience*, 1, 635–640.
- Blakemore S., J. Smith, R. Steel, E. C. Johnstone, & C. D. Frith (2000). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: Evidence for a breakdown in self-monitoring. *Psychological Medicine*, 30, 1131–1139.
- Botvinick, M. & J. Cohen (1998). Rubber hands ‘feel’ touch that eyes see. *Nature*, 391, 756.
- Daprati, E., N. Franck, N. Georgieff, J. Proust, E. Pacherie, J. Dalery, & M. Jeannerod (1997). Looking for the agent. An investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition*, 65, 71–86.
- Decety, J. & M. Jeannerod (1996). Fitts’ law in mentally simulated movements. *Behavioral Brain Research*, 72, 127–134.

- Decety, J., M. Jeannerod, & C. Prablanc (1989). The timing of mentally represented actions. *Behavioural Brain Research*, 34, 35–42.
- Decety, J., M. Jeannerod, D. Durozard, & G. Baverel (1993). Central activation of autonomic effectors during mental simulation of motor actions in man. *Journal of Physiology*, 461, 549–563.
- Dierks, T., D. E. J. Linden, M. Jandl, E. Formisano, R. Goebel, H. Lanferman, & W. Singer (1999). Activation of the Heschl's gyrus during auditory hallucinations. *Neuron*, 22, 615–621.
- Farné, A., F. Pavani, F. Meneghello, & E. Ladavas (2000). Left tactile extinction following visual stimulation of a rubber hand. *Brain*, 123, 2350–2360.
- Feinberg, I. (1978). Efference copy and corollary discharge. Implications for thinking and its disorders. *Schizophrenia Bulletin*, 4, 636–640.
- Fourneret, P. & M. Jeannerod (1998). Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia*, 36, 1133–1140.
- Frak, V. G., Y. Paulignan, & M. Jeannerod (2001). Orientation of the opposition axis in mentally simulated grasping. *Experimental Brain Research*, 136, 120–127.
- Franck, N., C. Farrer, N. Georgieff, M. Marie-Cardine, J. Daléry, T. D'Amato, & M. Jeannerod (2001). Defective recognition of one's own actions in schizophrenic patients. *American Journal of Psychiatry*, 158, 454–459.
- Frith, C. D., S. J. Blakemore, & D. M. Wolpert (2000). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society London B*, 355, 1771–1788.
- Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, 4, 14–21.
- Gallese, V. & A. Goldman (1998). Mirror neurons and the simulation theory of mind reading. *Trends in Cognitive Science*, 2, 493–501.
- Georgieff, N. & M. Jeannerod (1998). Beyond consciousness of external reality. A “Who” system for consciousness of action and self-consciousness. *Consciousness and Cognition*, 7, 465–477.
- Goldman-Rakic, P. S. & L. D. Selemon (1997). Functional and anatomical aspects of prefrontal pathology in schizophrenia. *Schizophrenia Bulletin*, 23, 437–458.
- Haggard, P., S. Clark, & J. Kalogeras (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, 5, 282–285.
- Harris, C. S. (1965). Perceptual adaptation to inverted, reversed and displaced vision. *Psychological Review*, 72, 419–444.
- Hesslow, G. (2002). Conscious thought as simulation of behavior and perception. *Trends in Cognitive Science*, 6, 242–247.
- Jeannerod, M. (1994). The representing brain. Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, 17, 187–245.
- Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia*, 33, 1419–1432.
- Jeannerod, M. (1999). To act or not to act: Perspectives on the representation of actions. *Quarterly Journal of Experimental Psychology*, 52A, 1–29.
- Jeannerod, M. (2001). Neural simulation of action: A unifying mechanism for motor cognition. *Neuroimage*, 14, S103–S109.

- Jeannerod, M. & V. G. Frak (1999). Mental simulation of action in human subjects. *Current Opinions in Neurobiology*, 9, 735–739.
- Johnson, S. H. (2000). Thinking ahead: The case for motor imagery in prospective judgments of prehension. *Cognition*, 74, 33–70.
- Knoblich, G. (2002). Self recognition: Body and action. *Trends in Cognitive Science*, 6, 447–449.
- Knoblich, G. & R. Flach (2001). Predicting the effects of actions: Interactions of perception and action. *Psychological Science*, 12, 467–472.
- Knoblich, G. & W. Prinz (2001). Recognition of self-generated actions from kinematic displays of drawing. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 456–465.
- Libet, B., C. A. Gleason, E. W. Wright, & D. K. Perl (1983). Time of conscious intention to act in relation to cerebral activities (readiness potential). The unconscious initiation of a freely voluntary act. *Brain*, 102, 193–224.
- Lipps, T. (1903). *Aesthetik: Psychologie des Schönen und der Kunst*. Hamburg: Voss.
- Nielsen, T. I. (1963). Volition: A new experimental approach. *Scandinavian Journal of Psychology*, 4, 225–230.
- Nielsen, T. I. (1978). *Acts. Analyses and syntheses of human acting, concerning the subject and from the standpoint of the subject*. Copenhagen: Dansk Psykologisk Forlag.
- Nielsen, T. I., N. Praetorius, & R. Kuschel (1965). Volitional aspects of voice performance: An experimental approach. *Scandinavian Journal of Psychology*, 6, 201–208.
- Paccalin, C. & M. Jeannerod (2000). Changes in breathing during observation of effortful actions. *Brain Research*, 862(1–2), 194–200.
- Parsons, L. M. (1994). Temporal and kinematic properties of motor behavior reflected in mentally simulated action. *Journal of Experimental Psychology. Human Perception and Performance*, 20, 709–730.
- Ramachandran, V. S. & D. Rogers-Ramachandran (1996). Synaesthesia in phantom limbs induced with mirrors. *Proceedings of the Royal Society London*, B263, 377–386.
- Schneider, K. (1955). *Klinische Psychopathologie*. Stuttgart: Thieme Verlag.
- Sirigu, A., J.-R. Duhamel, L. Cohen, B. Pillon, B. Dubois, & Y. Agid (1996). The mental representation of hand movements after parietal cortex damage. *Science*, 273, 1564–1568.
- Slachewsky, A., B. Pillon, P. Fournieret, Pradat-Diehl, M. Jeannerod, & B. Dubois (2001). Preserved adjustment but impaired awareness in a sensory-motor conflict following prefrontal lesions. *Journal of Cognitive Neuroscience*, 13, 332–340.
- Spence, S. A., D. J. Brooks, S. R. Hirsch, P. F. Liddle, J. Meehan, & P. M. Grasby (1997). A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of control). *Brain*, 120, 1997–2011.
- Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology*, 43, 482–489.
- Sullivan, R. (1969). Experimentally induced somatagnosia. *Archives of General Psychiatry*, 20, 71–77.
- van den Bos, E. & M. Jeannerod (2002). Sense of body and sense of action both contribute to self recognition. *Cognition*, 85, 177–187.

- von Holst, E. & H. Mittelstaedt (1950). Das Reafferenzprinzip. Wechselwirkungen zwischen Zentralnervensystem und Peripherie. *Naturwissenschaften*, 37, 464–476.
- Wegner, D. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.
- Weinberger, D. R. & K. F. Berman (1996). Prefrontal function in schizophrenia: Confounds and controversies. *Philosophical Transactions of the Royal Society, B*, 351, 1495–1503.
- Weinstein, E. A. & R. L. Kahn (1959). Symbolic reorganisation in brain injuries. In *American Handbook of Psychiatry*. New York: Basic Books.
- Wolpert, D. M., Z. Ghahramanin, & M. I. Jordan (1995). An internal model for sensorimotor integration. *Science*, 269, 1880–1882.

CHAPTER 5

Agency, ownership, and alien control in schizophrenia

Shaun Gallagher

University of Central Florida

One of the most puzzling things about certain positive symptoms of schizophrenia is the experience of alien control, that is, the sense that one is being controlled by other people or other things. This occurs in symptoms like delusions of control, where a schizophrenic person claims that someone else is causing him to act in a certain way or that someone else has his body under control. It also occurs in thought insertion, for example, when a person claims that someone else is projecting his thoughts into his mind. Recent cognitive and neuropsychological theories that have tried to explain just these kinds of schizophrenic symptoms never adequately address this specific issue. Explanations of how a person may come to lose the sense of agency for his own thoughts and actions fail to explain what seems to be the next step – the attribution of the thought or action to someone else. Some theories simply assume that a disruption in the sense of self-agency automatically leads to the misattribution of agency to someone or some thing else, or that the subject makes the misattribution as an inference from the fact that he is not the agent. But there are clear cases in normal experience when a person can lack a sense of agency for his action and yet not attribute agency to someone else. The most obvious example is the case of reflex movement. There must be, then, in the case of the schizophrenic symptom, some other explanation for the misattribution of agency.

In this chapter I will argue that recent top-down accounts of schizophrenic symptoms do not provide an adequate explanation for the misattribution of agency or the experience of alien control. In place of such top-down explanations, I develop the framework for a bottom-up account that shows how

the experience of alien control can be an implicit feature of the first-order phenomenal experience of some schizophrenic individuals.

1. Senses of agency and ownership

There are two sets of distinctions that are immediately relevant to our concerns. First, using a variety of methodologies, at least three different levels can be distinguished in the human cognitive system.

1. Non-conscious, sub-personal processes that are best described as neuronal or brain processes.
2. First-order phenomenal experience; the immediate experience that constitutes the “what it is like” or qualitative feel of consciousness. This level of experience may also involve an intentionality or directionality toward an object; it is *about* something.
3. Higher-order cognition, which supports the reflective ability to make introspective or attributive judgments about one’s first-order experience.

There may be some intermediate levels of description (syntactical or representational) that are also understood as non-conscious – I leave this question aside. Also, I do not deny that there may be unconscious mental states, but I also leave this idea aside.

Second, phenomenological distinctions can be made between the *sense of ownership* and the *sense of agency* as distinct aspects of self-awareness. In the normal phenomenology of voluntary action, it is difficult to see the distinction between the sense that this is *my* action (ownership) and the sense that this action is caused by me (agency). In the case of involuntary movement, however, these two things come apart. If someone pushes me from behind, I have a sense that it is my body that is moving, that the movement in question is in fact happening to me – I am the one who is moving. This involves the sense of ownership for that movement. But I also have the clear sense that I did not cause the movement, and hence I have no sense of agency for it.

Less intuitive is the idea that this distinction applies to thought processes as well. There is a sense of ownership and agency for our stream of consciousness. We may see this in the case of what Frankfurt (1996) calls “unbidden thoughts”. Thoughts may occur in my stream of consciousness – they are thoughts that I am having – they are not occurring in someone else’s stream of consciousness (thus I have a sense of ownership for them) – but I am not intentionally generating these thoughts (so I have no sense of agency for them). In such cases

I may be actively thinking of certain things and suddenly an unbidden memory is thrown into the mix, and I find myself thinking about things that are seemingly irrelevant, or that I do not want to think about. It may also be that someone else is causing me to think of certain things. This frequently happens in academic situations as I listen to a lecture and thus allow someone else to guide my thought. In such cases I may say that I do not experience a sense of agency for those thoughts, although I may rightly say that my agency is involved in actively listening to the other person.

The first set of distinctions between different levels of the cognitive system cuts across the second distinction between ownership and agency. Specifically, ownership and agency are manifested on at least two levels of self-conscious experience. At the first-order phenomenal level, the sense of ownership is the pre-reflectively experiential, felt, or lived-through sense that the thought occurs in *my* stream of consciousness or that the movement is of *my* body – phenomenologists refer to this as the *mineness* of experience. This is phenomenally distinct from the sense of agency, the felt sense that I am the cause or author of the thought or movement (Gallagher 2000a, 2000b). At the higher-order level of cognition, in the case of introspective self-consciousness, for example, one can also distinguish between:

- *Attributions of subjectivity* (or attributions of ownership) in which I reflectively attribute ownership to myself or report that “I am the subject in whom the thought is occurring” (or “It is *my* body that is moving”), and
- *Attributions of agency* in which I reflectively attribute agency to myself, or report that “I am the cause or author of the thought (movement).”

(Graham & Stephens 1994; Stephens & Graham 2000)

One important question is this: Which level of self-awareness has a generative (constitutional) priority? Does the second-order, introspective self-attribution generate a phenomenal sense of self-agency – does it, for example, operate as a kind of fiction-generating interpretation of first-order experience? Or does second-order self-attribution somehow depend on a genuine first-order sense of self-agency? As Mullins and Spence (2003:293) put it, “Whether thought insertion is solely an abnormal belief (or may also be an experience) is open to question.”

This distinction helps to distinguish between two different accounts of self-awareness: top-down accounts *versus* bottom-up accounts. It is also the case, however, that if we assume, roughly, that mental states are generated by brain states, we add a new level of complication to the question of generational priority. One can think of certain brain processes underpinning first-order phe-

nominal experiences, and other brain processes underpinning second-order reflective consciousness.

2. Top-down explanations

The distinctions between ownership and agency and the different explanatory strategies (top-down vs. bottom-up) have been put to work in explanations of positive symptoms of schizophrenia (including delusions of control, thought insertion, auditory hallucinations). I will argue that a bottom-up account, which takes first-order phenomenal experience seriously, provides a better framework for explanation of such symptoms, and that, based on scientific evidence, it is preferable to a top-down account.

Let me begin with some sample reports of delusions of control offered by patients and cited by researchers.

- (1) My grandfather hypnotized me and now he moves my foot up and down.
- (2) They inserted a computer in my brain. It makes me turn to the left or right. (Frith, Blakemore, & Wolpert 2000:358)

The structure of these reported experiences is similar to that of involuntary movement. My body moved (ownership), but I did not cause the movement (no agency) [*someone else caused the movement*].

Examples of thought insertion manifest the same kind of structure. “Simon” describes household appliances that are controlling his mind.

The things that come [to mind] are not the things that I have been thinking about ... They [the appliances] kind of short circuit the brain, and bring their message. (Jackson & Fulford 1997)

In terms of structure, there is a thought in my stream of consciousness (ownership), but I am not the author of that thought (no agency) [*someone or something else is causing me to think in this way*].

In all of these cases, the problem concerns a lack of a sense of self-agency rather than ownership, and in addition, misattribution to another cause. Subjects (1) acknowledge that the movements are happening to their own body; or thoughts are happening in their own stream of consciousness, and (2) they claim they are not the agents of these movements or thoughts, but rather, (3) they misattribute agency to someone or something else.

An adequate explanation of such symptoms would need to explain four things.

1. What is happening at the neurological level.
2. What is happening at the level of first-order phenomenal experience.
3. Why the subject acknowledges that she is the subject of, but not the agent of the movement or thought (why there is an attribution of ownership but not of agency on the higher-order reflective level).
4. Why the subject misattributes the agency to another person (or machine, or thing).

A top-down explanation of delusions of control and thought insertion has been developed by Graham and Stephens (1994; Stephens & Graham 2000). Following Dennett and Flanagan, Graham and Stephens propose an explanation of the normal attribution of agency in terms of “our proclivity for constructing self-referential narratives” which allow us to explain our behavior retrospectively: “such explanations amount to a sort of theory of the person’s agency or intentional psychology” (1994: 101). This view is based on a theory of mind (“theory theory”) approach that claims that we reflectively make sense of our actions in the context of a set of consistent beliefs and desires that constitute folk psychology.

[Normally] the subject’s sense of agency regarding her thoughts ... depends on her belief that these mental episodes are expressions of her intentional states. That is, whether the subject regards an episode of thinking occurring in her psychological history as something she does, as her mental action, depends on whether she finds its occurrence explicable in terms of her theory or story of her own underlying intentional states. (Graham & Stephens 1994: 102)

According to this view, our sense of agency depends on taking an intentional and theoretical stance toward ourselves – and the schizophrenic person seemingly gets her theory wrong. Graham and Stephens go on to explain that if a patient does or thinks something concerning which she has no beliefs or desires that would explain or rationalize the action or the thought, the movement or thought would appear to be something she does or thinks *without intention*. To make sense of it introspectively, the subject *infers* that she is not the agent; she constructs a narrative that involves misattribution to others. Thus, her lack of a sense of agency for her actions is the result of a mistaken conclusion that she has reached, and the sense of alien control is simply a delusional belief.

Thus, positive schizophrenic symptoms of alien control are *inferential mistakes* made on the basis of introspective or perceptual self-observation. “On our account, what is critical is that the subject find her thoughts inexplicable in terms of beliefs about her intentional states” (Graham & Stephens 1994: 105).

Graham and Stephens' top-down explanation clearly ignores first-level phenomenology and has nothing to say about neurological processes that may be involved. For them the conscious experiences of schizophrenic symptoms are not *lived through* in any original sense, but are determined by theoretical mistakes made at higher cognitive levels. In the case of inserted thought, for example, I experience *thought X* as seeming not to be the result of my thinking only *after* the fact, and as a result of some reflective verification process that has failed.

Such top-down accounts are not uncommon. Ralph Hoffman (1986) proposes an account that is similar in terms of the role attributed to introspection. According to him, verbal hallucinations in schizophrenia depend on the failure of a self-corrective, introspective judgment that normally corrects the default and automatic inference that ordinary unintended instances of inner speech are caused by something other than myself. The default mechanism defines non-intentional actions/thoughts as not caused by me. Normally, however, on some second-order level, there is a properly functioning "self-corrective process" that vetoes this mechanism in cases of my own actions/thoughts, and verifies that in fact such actions/thoughts are caused by me. In schizophrenia, this second-order cognitive process fails and leads to the misattribution of actions/thoughts to others.

In addition to pure top-down models, there are also hybrid models – accounts that combine a top-down account with a neurological explanation – or at least an explanation of cognitive mechanisms intended to be cashed out neurologically. For example, Frith (1992) argues that schizophrenic symptoms involving delusions of control are due to problems in metarepresentational self-monitoring. Although metarepresentation is described as a full-fledged introspective act of reflection, Frith explains what goes wrong in terms of sub-personal motor control mechanisms – the failure of efferent copy to reach a neurological comparator. The failure in the self-monitoring, metarepresentational process leads the subject to infer that his movements/thoughts are controlled/inserted by someone else. Campbell (1999) takes the Frithian model in a slightly different direction. The problem is not with the introspective level of self-monitoring – that is, introspection plays no causal role in such cases, and in fact its self-monitoring function must continue to work properly even for cases of inserted thought. "It is the match between the thought detected by *introspection*, and the content of the efferent copy picked up by the comparator, that is responsible for the sense of [agency] of the thought" (Campbell 1999:90). To be consistent, however, hybrid accounts like this should attempt to link the higher-order problems with introspection to neuronal processes re-

sponsible for higher-order cognition. Both Frith and Campbell talk about disruptions in metarepresentational introspection, but at the neuronal level they propose mechanisms directly tied to non-cognitive motor control of action, that is, mechanisms that are much more likely to generate problems directly with first-order phenomenal experience than with introspective report. The question that is rarely addressed in either the top-down or the hybrid accounts pertains to the role of first-order phenomenal experience.

3. A phenomenologically guided bottom-up account

Mullins and Spence (2003) suggest that thought insertion is not simply a matter of delusional belief since patients often describe experiential aspects of these events. The patients' experience seems more like an internal hallucination or feeling than like holding a theory about the thought. The patients describe inserted thoughts "in terms of bad impulses or unpleasant visual images ... internal voices." Inserted thoughts "feel different", or feel inserted at a specific point of entry into the head. "Hence thought insertion might not be solely a belief: in some (if not all) patients it can incorporate abnormalities of perception" (Mullins & Spence 2003: 295).¹

An alternative explanation to the top-down approach, which regards these symptoms as simply a form of mistaken belief or bad theory, is one that takes the first-order phenomenal level seriously. According to this phenomenologically guided bottom-up view, the explanation starts by asking what the schizophrenic person's experience is really like. Disruptions at the level of first-order phenomenal experience are then explained in terms of a neurological analysis that would identify the primary neural mechanisms underlying this experience. This involves something like a search for the sense of agency in the brain. Following current terminology, we might call this the search for the neural correlates of the sense of agency (the NCSAs). Despite the unsettled status of this strategy of searching for neural correlates of experience,² I will argue that a bottom-up approach that takes the phenomenology seriously may lead to a better understanding of what happens at the higher-order level of cognition and misattribution.

To take the phenomenology seriously means first to acknowledge that thought insertion and delusions of control have not always been clearly defined. The distinction between the sense of agency and the sense of ownership, if understood as a difference on the first-order phenomenal level, can help to clarify such concepts and to distinguish different phenomena. Thought inser-

tion, for example, is not the same thing as “influenced thinking”, or obsessional thinking. In the latter case, there is no loss of the sense of agency. In the case of influenced (“made” or passive) thinking, the patient still has some sense of agency, although he feels that he is being made to think certain thoughts by some external force (see Mullins & Spence 2003, for discussion). Second, any attempt to isolate something like a pure experience of thought insertion will involve an abstraction from the messy and complex experience of the schizophrenic person. The first-order phenomenology of some schizophrenic people includes a variety of problems across a number of dimensions, and can reflect anomalies in form and not just idiosyncratic content (Parnas & Zahavi 2002). Such anomalies are not just with regard to agency, ownership, self-awareness, and other-awareness, but also with regard to sensory, emotional, intentional, volitional, and temporal aspects. Patients can feel depressed, withdrawn, hostile, suspicious or sympathetic, reflective or non-reflective, etc. They may not be able to feel pleasure. They may experience cognitive interference with perception, memory, linguistic and semantic confusions, changes in normal experiences of optical, acoustic, olfactory, gustatory, tactile senses, depersonalization, derealization, etc. (see Klosterkötter 1988). These complications should put us on guard against any easy or neat analysis that would single out something like thought insertion or delusions of control as a cleanly defined phenomenon. At the same time, to make headway on explaining any of these problems, one does need to focus on one symptom, or one set of symptoms at a time, keeping the rest in the near background and allowing them to qualify the simplicity of one’s analysis. These provisos are meant to qualify the following account.

There are a number of ongoing research projects that could count as contributing to the search for NCSAs.³ Here I want to look at one theoretical model that is based on a promising and ongoing empirical research project. This model is usually referred to as the “Who” system (Georgieff & Jeannerod 1998). Overlapping areas of brain activation, or “shared representations” in the motor, premotor, and prefrontal cortexes, are evident for motor action, the observation of another’s motor action, and the imaginative enactment (conscious simulation) of my own and another’s motor action. That is, the same neuronal areas are activated when I engage in intentional action and when I see or imagine such action performed by another person. In itself, the model of shared representations suggests that there may be a confusion concerning who is doing the action. Jeannerod and his colleagues point out, however, that the distinction between my own action and the action of the other may depend on the *non-overlapping* areas of this neural matrix. So, the search for NCSAs

would involve finding those brain areas that are activated when I engage in intentional action but are not activated when I see or imagine someone else similarly engaged, and vice versa. More generally, the question is what brain areas are activated when I have a sense that I am in control of my action, in contrast to when I feel that someone else is controlling it. A number of recent studies have pursued this question.

Neuroimaging experiments by Farrer and Frith (2002), for example, have shown contrasting activation in the right inferior parietal cortex for perception of action caused by others, and in the anterior insula bilaterally when action is experienced as caused by oneself. Such activation is consistent with the idea that actions performed by others are perceptually mapped in allocentric coordinates (Jeannerod 1999). Farrer and Frith note that “there is strong physiological evidence that the inferior parietal cortex [involves this] kind of [allocentric mapping] ... to generate representations of body movements [by others] in allocentric coordinates” (2002:601). In contrast, the anterior insula involves information that is specified in egocentric spatial coordinates. Specifically, the anterior insula involves the integration of various kinds of self-specifying signals generated in self-movement: proprioceptive, visual and auditory ecological information about movement, and corollary discharge associated with motor commands that control movement. It is likely, as Farrer and Frith conclude, that a “close correspondence between all these signals helps to give us a sense of agency” (2002:602).⁴

Other studies have shown the relevance of these results for explaining the loss of a sense of agency in schizophrenic patients. In psychiatric and neurological patients, self-awareness disorders have been linked to metabolic abnormalities in the right inferior parietal cortex. In schizophrenic patients, specifically, the feeling of alien control (delusions of control) during a movement task has been associated with an increased activity in the right inferior parietal lobe (Spence et al. 1997).

It is still an open question, however, whether such models of motor control can be generalized to explain thought insertion, as Frith (1992) suggests (see Gallagher 2004). There may be other factors in the complex workings of the brain that contribute to a sense of agency and that apply equally to movement and cognition. One of these concerns timing, and specifically anticipatory aspects of movement and thought. Accordingly, problems with working memory may account for thought insertion in a more direct way. It has been shown that problems with working memory involving slight temporal differences are involved in schizophrenic subjects with delusions of control (Daprati et al. 1997; Franck et al. 2001; Vogeley et al. 1999). Similar problems with working mem-

ory, and specifically what phenomenologists call the protentional aspects of the stream of consciousness, could also disrupt structural features of thought processes (see Gallagher 2000b).

The main point is that the experiments cited above show that abnormal activation of certain primary brain areas responsible for the basic control of movement and/or timing can generate disruptions at the level of first-order phenomenal experience. The neurological problems occur in very basic, primary mechanisms for motor control, or working memory, and generate a first-order *sense that I am not the agent* of my movement or thought. And this is just the phenomenology reported by the patient: the movement or thought seems to appear *as if from nowhere* – or as not caused by the patient – and therefore as neither intended nor anticipated.

It follows that the subject will reflectively acknowledge that she is the subject of, but not the agent of, the movement or thought (attribution of ownership but not agency). At the higher-order level of introspective report, or reflective metarepresentation, the subject simply reports what she experiences at the first-order level: she self-ascribes ownership but denies agency for her actions or thoughts. Thus, in contrast to Graham and Stephens' top-down explanation, what happens at the higher-order level *is not a mistake, but a report of what the subject actually experiences*. The patient's introspection, interpretation, or narrative is correct, and reports a legitimate complaint: something is missing; a natural sense of self-agency. In other words, the higher-order cognitive level of attribution is an *explanandum* rather than an *explanans*, and is itself easily explained *from the bottom up*.

Indeed, this is the only way to make sense of the schizophrenic person's complaint about the lack of self-agency. Some level of the schizophrenic person's thought must remain unaffected by the disruption of the sense of agency, otherwise she would be unable to complain of alien movements, thoughts, or voices. If the higher-order introspective level was the problem, as Graham and Stephens suggest, then how would the report be generated as a complaint, as it usually is, that someone else seems to cause the movement or the thought. How would the very introspective thought itself escape this alienation? Why would schizophrenic people who suffer from delusions not feel equally alienated from their own complaints? Yet, the higher-order cognitive complaints are seemingly made in their own voice and with their own sense of agency intact.

4. Misattribution of agency

But this still leaves an important question unanswered. Why does the subject misattribute agency to another person (or machine, or thing)? That is, the schizophrenic person's complaint is not simply that he/she lacks a sense of agency for his/her actions and thoughts, but that someone else is actually controlling him/her.

There are two possible explanations. The first follows a commonly suggested top-down strategy. The cause of the misattribution is based on inferences made at the higher-order level of attributive consciousness. The subject, faced with the lack of a sense of agency for some thought or movement, introspectively interprets her experience as caused by someone else. The problem of misattribution is purely a higher-order introspective or narrative problem, a reasonable reaction to the loss of a sense of self-agency at the phenomenal level.

Another possibility, however, capitalizes on what we have already learned in the bottom-up approach. If it is in fact the case that some neurological component (the "Who" model) responsible for the differentiation between self and other has been disrupted, then it is quite possible that as a result, some sense of alterity is already implicit in the subject's first-order phenomenal experience. In this case, the attribution of agency to another is not the result of a theoretical stance that would read into first-order experience something that is not there. The latter would be a supplemental account generated in introspection; the odd result of a productive narrative. Rather, the bottom-up model suggests that the attribution of the action or thought to another is a genuine report of what is truly experienced in the pathology. The schizophrenic person who is experiencing delusions of control and thought insertion may be doing just that – *experiencing at the first-order level* that her actions/thoughts are from all appearances caused by someone else. Her attribution is not a mistaken inference or conclusion, and in *that* sense is not a "*mis*-attribution" at all. Rather, it is a correct report on her experience.

It may be possible to develop more evidence for this view. If the specific type of higher-order introspective cognition found in schizophrenic patients can also be found in other pathologies that do not involve introspective alienation, then just this kind of introspective cognition would not be sufficient to explain the schizophrenic effects (e.g., the misattribution of agency) manifested in thought insertion and delusions of control. In regard to characterizing the specific kind of introspection involved in schizophrenia, there are two possibilities, both discussed by Graham and Stephens.

First, using a Frithian view, metarepresentational introspection fails by inadequately monitoring the subject's actions – in effect, there is too little monitoring going on (Frith 1992). Second, according to a Sassian view, introspection goes wrong by becoming hyper-reflective – in effect, there is too much monitoring (Sass 1992, 1998, 2003).⁵ It turns out, however, that one can find pathologies other than schizophrenia that manifest one or the other of these abnormal forms of metarepresentation, but do not involve introspective alienation.

In some cases of utilization behavior, for example, subjects simply do not notice that their non-intentional and totally inappropriate actions are inappropriate. As in the Frithian failure of metarepresentation, the patient with utilization behavior fails to monitor his actions properly. In other cases the subject is aware of the inappropriateness of his action, experiences it as non-intentional, and may even try to resist it. Nonetheless, this kind of patient does not misattribute the agency of his action to someone else.

In contrast, in the case of obsessive-compulsive disorder, one finds precisely the kind of Sassian hyperreflection typical of some schizophrenics. People who suffer from obsessive-compulsive disorder do feel their actions or thoughts to be passively generated, unintended, and outside of their control. Yet, this hyper-reflection does not lead to a misattribution of agency. The obsessive-compulsive patient does not claim that someone else is causing him to do what he is doing. This is also the case with Anarchic Hand Syndrome (as distinguished from Alien Hand Syndrome); patients do not have a delusion that someone else is *actually* performing the action; rather, it is “*as if*” someone else were controlling their own arms, or simply, that there is something wrong with their arms (Frith & Gallagher 2002; see reports in Della Sala 2000; Marchetti & Della Sala 1998; Tow & Chua 1998).

These observations suggest that regardless of the nature of the higher-order cognition that exists in schizophrenia (whether it is *hypo* or *hyper*), it is not led, solely on its own (that is, by its own mistaken inferences or misinterpretations or corrupt theories) to misattributions of agency. Rather, it is likely that it discovers an alterity that is already manifesting itself on the first-order phenomenal level, generated by neuronal processes that are abnormal.

5. Conclusion

My intention here has not been to offer a complete explanation of these schizophrenic symptoms, and certainly not to explain the entire pathology of

schizophrenia – much more work would have to be done towards that end. Furthermore, the bottom-up account is not meant to deny the fact that odd, paradoxical, and wildly delusional narratives are often generated as the illness develops. The initial motivation for such narratives, however, may be shaped by processes that start out (for example, in prodromal phases of schizophrenia) as completely rational at the second-order level – that is, processes that generate a completely correct report of what the subject experiences.

I have argued, however, that, in contrast to the Graham and Stephens' account, schizophrenic delusions of control and thought insertion can be explained as disruptions of primary neurological processes that normally generate a first-order sense of self-agency. There are good logical and empirical reasons to think that, at least in the early stages of schizophrenia where clear reports of alien control are given, these reports are phenomenologically trustworthy, that is, trustworthy with respect to the subject's actual experience.

It is likely that more nuanced explanations are needed for more complex phenomena and that some combination of top-down and bottom-up explanation is needed to provide a more adequate account. It seems reasonable to assume that what C. Robert Cloninger (2002: 83) recently suggested about classification of mental disorders holds true for explanation of such disorders as well: that neither "mind-less nor brain-less approaches" will be adequate.

Notes

1. Frith (2003:240) agrees: "The patient with a delusion of control falsely believes that his actions are being directly controlled by alien forces. [This lies] so far from our normal experience that these disorders were typically classified as examples of false beliefs rather than abnormal experiences. Recent formulations suggest that these disorders are better characterised as abnormal experiences that derive their particular form from basic neural functions."
2. The search for something like an NCSA is somewhat abstract, misleading, and reductionistic unless we keep in mind that such neuronal processes are part of a larger system of embodied action in physical and social environments. For a debate about the very concept of searching for neural correlates, see Noë and Thompson (2004) and their commentators.
3. One such project involves identifying the elements of the forward model of motor control, and specifically the elements that contribute to the sense of agency (Blakemore, Oakley, & Frith 2003; Frith, Blakemore, & Wolpert 2000; Wolpert et al. 1998). I've discussed this model elsewhere, especially in terms of Frith's contribution to our understanding of how this model may explain delusions of control and its application to thought insertion (see Gallagher 2000a, 2004).

4. Studies by Farrer, Franck, Georgieff, Frith, Decety, and Jeannerod (2003) support this conclusion. It is important to note that in the experiments mentioned here, the authors have adopted the same concept I have defined above as the sense of agency, and as outlined in my 2000 paper in *Trends in Cognitive Sciences* (Gallagher 2000a). Other empirical studies consistent with the findings mentioned here have also used this definition (see, e.g. Blakemore et al. 2000; Chaminade & Decety 2002; Fournier et al. 2001; Jeannerod 2003; Ruby & Decety 2001; van den Bos & Jeannerod 2002; Vogeley et al. 2001; Vogeley & Fink 2003).
5. In Sass's bottom-up account, neurological problems may cause tacit sensory-motor processes, normally implicit in first-order phenomenal experience, to become abnormally explicit. This becoming explicit is already a form of automatic, or what Sass (2000) calls "operative", hyperreflexivity, and it may motivate more willful forms of hyper-reflective awareness.

References

- Blakemore, S. J., D. A. Oakley, & C. D. Frith (2003). Delusions of alien control in the normal brain. *Neuropsychologia*, 41, 1058–1067.
- Blakemore, S. J., J. Smith, R. Steel, et al. (2000). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: Evidence for a breakdown in self-monitoring. *Psychological Medicine*, 30(5), 1131–1139.
- Campbell, J. (1999). Immunity to error through misidentification and the meaning of a referring term. *Philosophical Topics*, 26, 89–104.
- Chaminade, T. & J. Decety (2002). Leader or follower? Involvement of the inferior parietal lobule in agency. *Neuroreport*, 13(1528), 1975–1978.
- Cloninger, C. R. (2002). Psychobiology and treatment of borderline personality disorder. *Acta Neuropsychiatrica*, 14(2), 60–85.
- Della Sala, S. (2000). Anarchic hand: The syndrome of disowned actions. *Creating Sparks, The BA Festival of Science*. www.creatingsparks.co.uk
- Daprati, E., N. Franck, N. Georgieff, J. Proust, E. Pacherie, J. Dalery, & M. Jeannerod (1997). Looking for the agent: An investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition*, 65, 71–96.
- Farrer, C., N. Franck, N. Georgieff, C. D. Frith, J. Decety, & M. Jeannerod (2003). Modulating the experience of agency: A positron emission tomography study. *Neuroimage*, 18(2), 324–333.
- Farrer, C. & C. D. Frith (2001). Experiencing oneself vs. another person as being the cause of an action: The neural correlates of the experience of agency. *NeuroImage*, 15, 596–603.
- Fournier, P., N. Franck, A. Slachevsky, et al. (2001). Self-monitoring in schizophrenia revisited. *Neuroreport*, 12(6), 1203–1208.
- Fournier, P. & M. Jeannerod (1998). Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia*, 36, 1133–1140.

- Franck, N., C. Farrer, N. Georgieff, M. Marie-Cardine, J. Daléry, d T. 'Amato, & M. Jeannerod (2001). Defective recognition of one's own actions in patients with schizophrenia. *American Journal of Psychiatry*, 158, 454–459.
- Frankfurt, H. (1976). Identification and externality. In A. O. Rorty (Ed.), *The identities of persons* (pp. 239–251). Berkeley: University of California Press.
- Frith, C. D. (1992). *The cognitive neuropsychology of schizophrenia*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Frith, C. D. (2003). The pathology of experience. *Brain*, 127(2), 239–242.
- Frith, C. D., S.-J. Blakemore, & D. M. Wolpert (2000). Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain Research Reviews*, 31(2–3), 357–363.
- Frith, C. & S. Gallagher (2002). Models of the pathological mind. *Journal of Consciousness Studies*, 9(4), 57–80.
- Gallagher, S. (2000a). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, 4(1), 14–21.
- Gallagher, S. (2000b). Self-reference and schizophrenia: A cognitive model of immunity to error through misidentification. In D. Zahavi (Ed.), *Exploring the self: philosophical and psychopathological perspectives on self-experience* (pp. 203–239). Amsterdam & Philadelphia: John Benjamins.
- Gallagher, S. (2003). Schizophrenia and the narrative self. In A. S. David & T. Kircher (Eds.), *The self and schizophrenia: A neuropsychological perspective* (pp. 336–357). Cambridge: Cambridge University Press.
- Gallagher, S. (2004). Neurocognitive models of schizophrenia: A neurophenomenological critique. *Psychopathology*, 467, Vol. 37, 8–19.
- Georgieff, N. & M. Jeannerod (1998). Beyond consciousness of external events: A Who system for consciousness of action and self-consciousness. *Consciousness and Cognition*, 7, 465–477.
- Graham, G. & G. L. Stephens (1994). Mind and mine. In G. Graham & G. L. Stephens (Eds.), *Philosophical psychopathology* (pp. 91–109). Cambridge, MA: MIT Press.
- Hoffman, R. (1986). Verbal hallucinations and language production processes in schizophrenia. *Behavioral and Brain Sciences*, 9, 503–517.
- Jackson, M. C. & K. W. M. Fulford (1997). Spiritual experience and psychopathology. *Philosophy, Psychiatry, and Psychology*, 4, 41–65.
- Jeannerod, M. (1999). To act or not to act: Perspectives on the representation of actions. *Quarterly Journal of Experimental Psychology, A* 52, 1–29.
- Jeannerod, M. (2001). Neural simulation of action: A unifying mechanism for motor cognition. *Neuroimage*, 14, S103–S109.
- Jeannerod, M. (2003). The mechanism of self-recognition in humans. *Behavioral and Brain Research*, 142(1–2), 1–15.
- Jeannerod, M., C. Farrer, N. Franck, P. Fournier, A. Posada, E. Daprati, & N. Georgieff (2003). Action recognition in normal and schizophrenic subjects. In: T. Kircher & A. David (Eds.), *The self in schizophrenia: A neuropsychological perspective*. Cambridge: Cambridge University Press.
- Jeannerod, M. & V. Frak (1999). Mental imaging of motor activity in humans. *Current Opinions in Neurobiology*, 9, 735–739.

- Klosterkötter, J. (1988). *Basissymptome und Endphänomene der Schizophrenie*. Berlin: Springer.
- Marchetti, C. & S. Della Sala (1998). Disentangling the alien and anarchic hand. *Cognitive Neuropsychiatry*, 3(3), 191–207.
- Mellor, C. S. (1970). First rank symptoms of schizophrenia. *British J Psychiatry*, 117, 15–23.
- Mullins, S. & S. A. Spence (2003). Re-examining thought insertion: Semi-structured literature review and conceptual analysis. *British Journal of Psychiatry*, 182, 293–298.
- Noë, A. & E. Thompson (2004). Are there neural correlates of consciousness? *Journal of Consciousness Studies*, 11(1), 3–28.
- Parnas, J. & D. Zahavi (2002). The role of phenomenology in psychiatric diagnosis and classification. In M. Maj et al. (Ed.), *Psychiatric diagnosis and classification* (pp. 137–162). Chichester: John Wiley and Sons.
- Ruby, P. & J. Decety (2001). Effect of subjective perspective taking during simulation of action: A PET investigation of agency. *Nature Neuroscience*, 4(5), 546–550.
- Sass, L. (1992). *Madness and modernism: Insanity in the light of modern art, literature, and thought*. New York: Basic Books.
- Sass, L. (1998). Schizophrenia, self-consciousness and the modern mind. *Journal of Consciousness Studies*, 5, 543–565.
- Sass, L. (2003). Schizophrenia and the self: Hyper-reflexivity and diminished self-affection. In T. Kircher & A. David (Eds.), *The self in schizophrenia: A neuropsychological perspective* (pp. 242–271). Cambridge: Cambridge University Press.
- Spence, S. A., D. J. Brooks, S. R. Hirsch, P. F. Liddle, J. Meehan, & P. M. Grasby (1997). A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control). *Brain*, 120, 1997–2011.
- Stephens, G. L. & G. Graham (2000). *When self-consciousness breaks: Alien voices and inserted thoughts*. Cambridge, MA: MIT Press.
- Tow, A. M. & H. C. Chua (1998). The alien hand sign—case report and review of the literature. *Ann-Acad-Med-Singapore*, 27(4), 582–585.
- van den Bos, E. & M. Jeannerod (2002). Sense of body and sense of action both contribute to self-recognition. *Cognition*, 85(2), 177–187.
- Vogele, K., M. Kurthen, P. Falkai, & W. Maier (1999). The human self construct and prefrontal cortex in schizophrenia. The Association for the Scientific Study of Consciousness: Electronic Seminar (<http://www.phil.vt.edu/assc/esem.html>).
- Vogele, K., P. Bussfeld, A. Newen, et al. (2001). Mind reading: Neural mechanisms of theory of mind and self-perspective. *Neuroimage*, 14(1), 170–181.
- Vogele, K. & G. R. Fink (2003). Neural correlates of the first-person-perspective. *Trends in Cognitive Sciences*, 7(1), 38–42.
- Wolpert, D. M., R. C. Miall, & M. Kawato (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, 2, 338–347.

Tetraplegia and self-consciousness

Jonathan Cole

Universities of Southampton and Bournemouth

1. Introduction: What self are we conscious of?

It is, arguably, a little rare to be conscious of “one’s own identity, actions and sensation” (as self-consciousness is defined by *the Oxford English Dictionary*) – these tend to be in the background (Gallagher 1986). This lack of self-consciousness in everyday life may be implicit in the other use of the term, “undue or morbid preoccupation with one’s own personality”; there is normally a fine line between being aware of self at all and being *too* aware.

One also has to consider which parts of the self one is conscious of, or attending towards. A mathematician may measure himself by the functioning of a small part of the frontal cortex, a soccer player by her cultured left foot. For many engaged in the analysis of consciousness, the self resides in the brain and in our ideas, thoughts, sensations, and feelings. The body, often phenomenologically absent (Leder 1990), can also be absent from considerations of the self. Within such a framework, neuropsychology has allowed insights into the fragility of the brain and the ways in which abnormalities of its function illuminate the mechanisms of mind (for examples see Weiskrantz 1997; Boks 2003).

For others, self-consciousness reaches out to the body and, indeed, to the world (see, for instance, Damasio 2000; Varela et al. 1991). Moreover, a cognitive-centred perception of self may not correspond with the predominant folk psychology view, which includes within one’s self the body in appearance and action. Work within neurology has shown the importance of various losses in movement and sensation in revealing function, but also the way in which such losses reveal some of the ways in which people view the self as being embodied (e.g., Sacks 1985; Cole 1995).

This paper will explore some of the consequences of a severe neurological impairment, spinal cord injury (SCI), which leads to a complete absence of movement and sensation below the neck (tetraplegia¹). In an instant, people lose bodily sensation or movement for the rest of their lives. Murphy (1987), an anthropologist who became tetraplegic, wrote, “My past is divided into two parts ... no longer smooth or linear, my history is bisected and polarised” Such a huge transition tests and threatens normal assumptions about the body and the self.

In exploring how people adapt their minds to their new bodies, the problems divide between the effects of the impairment itself on the body, “biological embodiment” (Shoemaker 1976), the problems in living in the world with such an impairment, “engagement with the world”, and the responses of others to the impairment and one’s new way of living, “social currency.” I will seek to show how a wide view of the impairment is necessary to understand its experience and – by reflection – to understand our normal, embodied existence. As Simon Brissenden (1990) wrote,

The experience of spinal injury cannot be understood in terms of purely internal psychological or interpersonal processes, but requires a whole range of other material factors such as housing, finance, employment, the built environment and family circumstances to be taken into account.

Those with SCI do not talk of self-consciousness or of self – and these terms will largely be absent from this chapter – but from their experience it is clear that their injuries have profound effects on the ways in which they view themselves. It is also clear that to regain self-respect most have to go beyond a consideration of a brain-focused self to a restoration of an embodied self in both the physical world and in society. The examples are taken from a book exploring these issues from the subject’s perspective (*Still Lives*, Cole 2004).

2. The neurological impairment; levels of loss

The neurological impairments following damage to the spinal cord include losses of movement and sensation, and disorders of autonomic control of bladder, bowel, and the circulation. The level of injury is crucial; those injured high in the neck will need assistance with breathing, those injured lower in the neck will be able to transfer from chair to bed etc., and may live independently. After thoracic injuries and below, people lose the use of the legs. The control of bladder and bowel is coordinated in the lower spinal cord lying just above the

sacrum, so anyone with a complete lesion above this will be incontinent. Men usually have either an indwelling urinary catheter or introduce a catheter intermittently (without sensation of course). Women usually need the former to prevent incontinence. The bowel can be trained to empty fairly regularly, or be evacuated manually, though continence is never secure. Sexual function is coordinated at similar levels to the bladder.

Most people with spinal cord injury also have to live with a variety of pain. Around 60–65% of people have pain and in 20–25% it is severe. This may reflect damage to the nerve roots at the site of the lesion, or may be due to shoulder arthritis after years of transfer from chair to car etc. One of the most troubling of pains is “phantom” pain perceived in the area of the body below the level of cord damage in area which cannot be felt.

3. The self and the impairment

3.1 The servant and the master

Though tetraplegia is a single medical impairment, the responses of injured people are very different. One man described how:

The first thing I assessed was what I could no longer do. What the future did not hold for me. I will not raise a family; I will not be independent. You are totally self-obsessed at the beginning. Someone else will have to do my bowel care, put me to bed and wash me.

For another, thirty years since his tetraplegia, his memory had become no less raw:

You can't imagine how devastating this was when you've just turned 20. And that they can't do anything about it. Absolutely nothing. You cannot imagine the anger, the grief, and the devastation.

The perceived self changes entirely. One man, embedded in his previous physicality, told me,

You just cannot substitute for the experience of being able to use this wonderful piece of equipment, the body, be it running, riding or shagging. My greatest passion was horse riding. The sheer enjoyment and freedom of being able to go hell for leather across the forest on the back of this living being with communication and some measure of control was awe-inspiring, wonderful and something that I cannot – ever – experience again.

He described, as others have done, that the awareness of his impairment never leaves and so in a way it is a disease of consciousness, inhabiting and inhibiting his self-perception and his being, “I can never forget that I am in a chair. I am devastated by my body.”

And yet for others, in a similar position, their tetraplegia – though a huge interruption in their lives – was less overwhelming. A man, injured soon after qualifying at university, told me:

I was concerned with the practicalities of getting on with life. I didn’t lie there thinking all the time “Oh My God what have I done?” I never burst into tears because, from the early stages of living with the injury, I saw the whole thing as a challenge. How do I overcome so and so? How do I deal with this? How do I come to terms with that? I never thought, “I can’t do that.”

Without an animated body, people are forced to re-examine their perception of themselves and one might expect them to become more intellectual and less physical. Yet, paradoxically, they have to attend to the body more. What before was automatic now has to be thought about. Previously automatic, autonomic activities, like bladder and bowel, have to be attended to. Without care of skin areas and the bladder, potentially fatal infection will set in. The body, previously a servant of the will, now becomes master. What was once often the pride of young men, now becomes lifeless, yet forces attention towards itself.

3.2 A voyage of discovery

One should not overlook the difficulties in learning to care for an unmoving body. Just sitting in a wheelchair without feeling from the head down, and without postural control, is very difficult. To learn new motor skills as an adult is difficult. One person, newly tetraplegic, described learning to catheterise himself,

I open a bit of sticky, get the catheter, shuffle forward on the chair, undo trousers and get everything out, connect, pass catheter and place over the loo. I do not know whether it [cognitive difficulty] is part of spinal cord injury but it has taken me 6 months to do this simply. Bowel care is the same, a voyage of discovery. After two weeks, fourteen times, you would think you’d know what to do, but I would still have accidents. Now, six months later, I am so comfortable with it, [but] it took me months.

3.3 The feeling of nothing

The requirement to look after the body is absolute. But what of the sensation of numbness, what might nothing feel like? I asked several tetraplegics, who naturally found it difficult to explain. One told me,

This is not the same sort of sensation somehow as having a tooth filled. That numbness is sort of a void, as though something has been taken away. My loss of sensation is not quite like that; it is a different feeling. The sensation of nothing was a sensation. It wasn't numbness, it was nothingness. It was a sensation because you can sense nothing.

The first time in a wheelchair after weeks lying down is often unnerving, with the person feeling a long way up as he looks down on his body, unfelt, unsafe, and precarious. "My head floating ... I felt like a balloon being wafted around. This was odd and un-nerving."

One woman told me that she localised her self to her sentient head and to her thoughts and ideas, and that she did feel disembodied. But most people adapt to become at one with their "new body". One man said,

It took weeks and months to feel OK. Now I can almost kid myself that I can feel something when I sit in a chair, even though I know I cannot. It feels exactly the same sitting in a chair now to before I was injured. It can't but it does. My mind tells me so. My mind makes me think I am like you over there. It learns what is the norm for this body. It tells me there is nothing wrong, so I feel comfortable and correct.

Another suggested that "The absence of sensation does not feel wrong. It seems perfectly normal." This may be the sort of restitution of normality that Goldstein (1939) discussed. This is not to suggest that the mind does not seek the body, or that tetraplegics do not try to connect with their bodies.

I cannot do anything myself to get a connection. And despite that absence I have to be aware of everything, since, for instance, if I reach out, my balance is so bad that I might fall. It is not natural so I am aware the whole time. If I pinch my legs it is numb. The pain is the connection – my friend the pain.

Several people told me that a small amount of pain from their bodies was good, since it allowed them some feeling. Alas even this is illusory, since the pain arises from the central nervous system and not from the body itself. But the preferment of pain to nothing shows the mind's need for the body.

Tetraplegia imposes on people and on their self-consciousness with its needs and by the limitations of movement and spontaneity it imposes. The first relationship after the onset of this devastating impairment is, necessarily, with

their bodies. But just as mathematicians and footballers have different views of their embodiment, so too do people after SCI. Some grieve for a time and carry on, some never get over it, but many, perhaps most, do reach an accommodation and learn to carry on. This does not seem to reflect their pre-impairment physicality – one gymnast, once injured, was able to move on without a backward glance, whilst a rugby player could not. Most people still feel it is their body and that they are, in some way, within it. Merleau-Ponty (1962) wrote that “I observe objects, but not my own body”. To an extent, tetraplegics do observe their own bodies, but, though they are not in the flesh fully, their selfhood reaches out to inhabit what they no longer can move or feel.

4. The world and the disability

4.1 Leaving stones outside?

Thus far we have concentrated on the relationship with the body. But tetraplegics still have to be concerned about moving through the world. Problems in access may prevent their taking part fully in society. Murphy (1987) quotes Mead, “an individual’s concept of his or her self is a reflection or, more accurately, a refraction, of the way he or she is treated by others.” Lack of accessible transport or curb cuts in the street not only prevent movement in a wheelchair but also suggest society does not care. Two people in the UK who have explored the gap between those with SCI and the world are Michael Oliver and Stephen Duckworth. Both tetraplegic, both active in the disability movement, their experiences have driven their views and their desire for social change.

Stephen Duckworth was a medical student when, playing rugby, he made a bad tackle and became tetraplegic. After a protracted rehabilitation period, complicated by life threatening infections, he returned to complete his medical studies from a wheelchair. Before he had been a rather typical rugby playing, sailing, medical student, eager to learn about diseases and anatomy, but far less sure talking with patients, especially those from the inner city. After two years of illness, he had learnt to listen.

Before I felt incompetent at interacting with other patients. [After] I felt brilliant at doing it. I was so proud of how I had changed as a person. I could spend hours talking to people who were dying, before I would never have gone there, it was too scary.

Despite the setbacks, he qualified. But then, before starting his pre-qualification house jobs, with a few months to fill, he went to a local Day Centre for “the disabled”. It was the first time Stephen had come across other types of disabled people, and it was revelatory.

My eyes were opened. I started to realise the indignity with which society treats disabled people. Society responded to disabled people by excluding them. Go to a Day Centre for disabled people and don’t expect to go to University, don’t expect to get a job, don’t have high expectations and certainly have low self-esteem. They had segregated transport so they went on the dial-a-ride bus instead of public transport. These were people with cerebral palsy, juvenile onset rheumatoid arthritis, mental health problems, difficulties with sight, grouped together by their exclusion from the mainstream.

Stephen decided that, instead of sweating through house jobs, he would take an MSc course at Southampton in Rehabilitation Studies. There he fell in with others with physical impairments, as well as other marginalized groups at the time, gays and blacks. His course required him to write a thesis and he chose, “The impact of medical education on the attitudes of medical students towards disabled people”.

As he had suspected, he found more sympathy and concern for the disabled in first-year medical students than there was in students in their final year – the demands of the course squeezed out their humanity. After his MSc, he read for a PhD on the problems those with disability have with employment. He posted his chapters off to a Member of Parliament and found himself being quoted in Parliament. He set up his own company to assist employers in hiring disabled people and to give seminars for the disabled to help them into work. He encouraged the disabled to take over their lives.

What we tried, and try, to do is to conceptualise their disability as being like a stone or stones (since many people have more than one problem). It might be a problem caused by disability, but equally by work, or a relationship with partner, whatever. We get them to quantify their stone and then ask if they can put it down. If they go to the cinema, then can they leave their stone outside?

Over the last twenty years, despite being able to move one arm and hand (poorly), Stephen has become a successful head of a company helping disabled people back into work (and so making them feel better about themselves) and lobbying for better awareness of the needs of the disabled both in work and in access. To do this he had to leave medicine, and not simply because it would have been physically demanding.

Within medicine, he had to accept the medical model of his injury; that it was tragic, catastrophic, and permanent, perhaps the ultimate loss – of movement, of sensation, of personal care, and of independence. He was expected to be looked after, to be an object, and to be passive. He explored other ways of looking at his condition and himself. To progress it was imperative that he did leave medicine, for medicine was then capable of viewing him and his condition in only one, pitiable, way.

4.2 “The best thing. . .”

After his SCI, Michael Oliver spent two years as a teenager at home being looked after by his parents. He raged, but not for the reasons one might have imagined.

I was angry when I was discharged not because I had not come to terms with my spinal injury, as I am sure my social worker would have written in her case notes, but because I had to eat, sleep and defecate in the same room, and that was the room I shared with my parents when we watched television as well.

Born in a far less privileged environment than was Stephen, Michael was rescued by a chance opportunity to work in a local remand home for adolescents. From there, though previously completely uninterested in education, he developed a passion for sociology and, after leaving school with no qualifications, is now a professor of disability studies. He has argued against many prevailing ideas including that one goes through stages of adjustment to neurological or other impairments, quoting Zola (1982):

I realise how meagre are our attempts to write about adjustment and adaptation. It would be nice if, at some point, growing up ends and maturity begins, or if one could say that successful adjustments and adaptations to a particular difficulty have been achieved. For most problems, or perhaps most basic life issues, there is no single time for such a resolution . . . The problems must be faced, evaluated, re-defined, and readapted to, again and again. I knew that this applied to me. No matter how much I was admired by others or by myself, there was still much more I had to face. “My Polio” and “My Accident” were not just my past: they were part of my present and my future.”

Oliver does not deny that there are periods of mourning and denial, etc., but suggests the stages of coming to terms with SCI are found rarely and often are designed to allow others to move on from the impairment. He could not have predicted his own response to SCI.

Once it did happen to me, all the things I thought I would think and feel, I never felt at all; we are not able truly to know ourselves, however well we think we do. If someone dies, or we go home and our partner says they are divorcing us, or we lose our job, though we might have thought about those things before and how we might think, I do not think any of us really knows how we would feel beforehand.

Mike Oliver and many others within the disabled movement have tried to change not only our perceptions about neurological impairment, but also the relations between those with SCI and their medical attendants. Vic Finkelstein, for instance, also living with spinal cord injury, has railed against the injustice and wrong headedness of his own rehabilitation:

The aim of returning the individual to normality is the central foundation stone of rehabilitation. If, as happened to me following my spinal injury, the disability cannot be cured, normative assumptions are not abandoned. On the contrary, they are re-formulated so that they not only dominate the treatment phase and totally colour the helper's perception of the rest of that person's life. The rehabilitation aim now becomes to assist the individual to be "as normal as possible".

The result, for me, were endless soul-destroying hours [in hospital] trying to approximate to able-bodied standards by "walking" with callipers and crutches ... Rehabilitation philosophy emphasises physical normality and, with that, the attainment of skills that allow the individual to approximate as closely as possible to able-bodied behaviour (e.g. only using a wheelchair as a last resort, rather than seeing it as disabled peoples' mobility aid like a pair of shoes is an able-bodied person's mobility aid). (Finkelstein 1988)

In trying to encourage an acceptance of difference and to help people see beyond the impairment, Oliver has challenged established views. His view of his SCI was expressed in a newspaper article. It was, he told me,

About not worse, not different – there is a third difference – that SCI makes life better. That is why I wrote the piece in "The Guardian" about my disability being the best thing that had even happened to me. Because – for me – I think it was. I was a working class yobbo with a failed education, not very good at relationships, in a job that I did not like and I probably would have gone on to drink [and smoke] too much.

Breaking my neck broke that mould and gave me an alternative possibility. It changed the possibility of who I could become. Forty years later I am a professor of disability studies, I have one marriage behind me and I am happily married again. I have grandchildren and have been all over the world. I have had a good life. I have no complaints. One thing I do know that if I had not

broken my neck I would not be a professor in a university. But equally well I do not want me to be positioned as some sort of hero. I have just taken opportunities as they occurred.

4.3 Models and definitions

Mike and many others in the “disability movement” have tried to redefine the terms and, more importantly, the perceptions of disability. If disability is seen as a tragedy, then disabled people will be treated as passive victims, and this will not only affect social policy but also the way they see themselves. Mike and others suggest another model. Their ideas have become more and more accepted and form the underpinning of an important document from the Royal College of Physicians of London and The Prince of Wales’ Advisory Group on Disability (1992). In this, the following definitions are given:

Impairment is the loss or abnormality of a particular faculty or part of the body.

Someone with a disabling impairment is a disabled person.

Disability refers to a disabled person’s encounter with daily living, the environment and society, not only in particular circumstances, but encompassing the whole of that experience.

This is all presaged by Oliver (1990):

The individual model [of disability] sees the problem as stemming from the functional limitations or psychological losses assumed to arise from disability. Nothing could be further from the truth. The social model suggests it is not the disability, not individual limitations, which are the cause of the problem but society’s failure to provide appropriate services and failure to ensure the needs of disabled people are fully taken into account. Hence disability is all the things which impose restrictions on disabled people, from individual prejudice to institutional discrimination, from inaccessible buildings to unusable transport systems.

Whilst others have maintained that one cannot rule out the limitations of the impairment from consideration entirely – and there has been a spirited debate about this within the disability movement for years – what Oliver, Duckworth, and others have done is to suggest new ways of looking at impairments, and by removing stigma about difference, pleaded for inclusion in work and accessibility. Thus, they suggest, it is from these actions by society that those with a disability can be made to feel better about themselves. The medical impair-

ment, they are saying, is manageable. It is the problems heaped on disabled people by society in terms of access and employment that undermine their self-respect.

The following is a well-known quote by Merleau-Ponty (1962): “The body is the general medium for having a world.” That is usually considered in terms of the body, but to the same degree, each body requires a world through which it can move, and to deny this is to negate the person and reduce his selfhood. As Murphy suggested, “Disability is not simply a physical affair for us: it is our ontology, a condition of our being in the world” (Murphy 1987). Any improvement in access for the disabled has profound effects on their self-esteem.

5. The other; social currency with SCI

5.1 Seeing the wheelchair?

It is one thing to feel like a member of society, but society is made up of individuals and individual relationships. Those with SCI have to re-establish such relationships to feel fully restored.

Julie Hill became paraplegic in a road traffic incident. She became famous for having an electrical implant that gave her movement in her legs, after a fashion, from stimulation of the spinal roots. One suspects it was this that enabled her to publish her autobiography, “Footprints in the Snow” (2000). But the book is more remarkable for the honesty with which she portrays her experiences than for the research. In particular, it details strains between her and her husband, Kevin.

After the injury, as she learnt to look after her new self, with new friends and with a new identity, Kevin, in contrast, was on his own and survived by adopting a routine with their two boys and their school, his work, and with cooking and shopping. His routine became so important that when, after several months, Julie started going home for visits she disrupted it:

When she came back for the weekend it was as though she was a visitor. I would clean the house in a certain way each time. That was very important to me; when she came back she messed up and interrupted this. It was nice to see her but I was pleased at the end of a weekend, once she had gone.

Julie felt the antagonism and blamed Kevin for seeing the wheelchair, not her. But for Kevin the situation was very different.

She thought the problem was my seeing the disability and not her. It was not. It was because all her friends and her new life were in the hospital and in the unit, and I could not be part of that. I was ostracised because of her rehabilitation.

For Kevin, the problem was that her relationship with the hospital, and the people there, distanced her from him and from life outside. At her 29th birthday celebration, there was family, and those with spinal cord injury and the nurses; two groups at opposite ends of the room all night. Initially Julie was at a very low ebb and needed help to gain a sense of her self as valid and useful; the hospital and her new friends helped her with this. But she needed also to become outward looking again, saying, “Yes I am paraplegic, but am I more than that.”

These were difficult times, but after a few months, Julie managed to tempt Kevin away for a holiday with just the two of them; they learnt to laugh together again, and over the years they have become closer than ever. But their experience exemplifies the way in which spinal cord injury may pull people apart. In struggling to come to terms with its effects on them, either as patient or partner, it reduces the ability to empathise with the other. The changes in self-consciousness after injury tilt the balances between consciousness of oneself and of others.

5.2 There is no manual

For Julie to feel good about her new self, she had to find something in her immobile and insentient body she could feel pleased with. In this, as in so much else, she succeeded and she even lectures on sex and SCI. But sexuality remains a large problem after spinal injury, since sexual identity is so difficult to divorce from the body.

Initially after SCI, thoughts of becoming sexually active again are often far away, despite an unaltered libido (remember many are fit young men). For some, a sexual identity never returns, but for others it does and so reveals something of the ingenuity and creativity with which people cope with their injury. Anne Seaman, who lectures on sexuality in SCI, told me, “People have to find out for themselves. They do understand their own bodies far more than normal people. There is no manual.”

For some, orgasm may come from other places and in other ways than before. There are many anecdotal reports of areas above the injury becoming remapped as an area for such gratification. She remembers touching one

man with tetraplegia on the neck during a conversation. He said simply, but unmistakably, "I save that."

Figures from marriages suggest that this relationship is no more or less successful in those with SCI than with others. In one sample, 66 people were single at the time of their injuries and 20 had married since their injuries. Of those 102 who were married at the time, 73 still were several years later, with 12 widowed (Morris 1991). Figures from the US suggest that divorce is 2.3 times more common after the first three years of injury, with stressors involving reduced ability to take care of a home, problems with work, childcare, mobility and finance, as well as psychological adjustment and sexual functioning, problems we all face. Among the single, there is a reduced rate of marriage in the first three years, but after that the rate increases and marriages after injury do rather better than those before (Stover et al. 1995). Anne Seaman told me that after SCI, people tend to do one thing per year for a while, either adapting their house, or having their first holiday and maybe being available for a relationship. The pace may have slowed, but their needs are as before and their success in the most intimate and enduring of relationships reflects the ways in which social competency can return after injury.

5.3 Leaving dependency

Up to the 60s and 70s, tetraplegics were either looked after by family or in "homes" for the "young disabled". Since then, increasing numbers have been able to live in their own homes independently, a huge advance on the regimented passivity of a care home. But, for this most tetraplegics need personal assistants (PA) who wash and bath their employers, as well as assisting with dressing, bladder and bowel care, and a range of other things. More than all these is the way in which a PA allows a transition from dependency to spontaneity, something as important for those who live with their loved ones as for those who live alone; people do not want their wives or husbands to assist with transfers, bladder drainage, etc., as well as be companion and lover.

Stephen Duckworth once talked of doing the washing up for his wife. He related this as his action, even though he did not do it physically. By asking his PA to do it, he gained a sense of agency and ownership of action. By acting and doing when the tetraplegic asks and in a manner satisfactory for him, a PA relinquishes her own agency and gives it to the tetraplegic, allowing him a real sense of acting in and on the world. This independence has been of great importance in returning self-esteem to those with SCI. The relationship with a PA is extraordinary. It:

Reverberates in every corner of one's life. It shares intimacy, it oversees success and failure. It is an interaction that is virtually unique. By its very nature it is physical and breaches strong social taboos. It is a relationship that is under constant flux: mutating as individual replaces individual. I have developed many short-term relationships. Each one of these ... is fully rounded and individual, capturing a period, a space, a moment... (from Vasey 1999)

The distance between employer and PA needs to be right. Beverly Ashton wrote:

Over the years I've caused myself a lot of difficulty by listening to, worrying about and taking on the life problems of PAs. I now know, without a shadow of doubt, that the best response when somebody tearfully spills out some terrible problem, is to say "Oh Dear! I'm so sorry. That must be awful ... but could you look to see whether the vacuum cleaner bag needs emptying.

I have been repeatedly involved in really desperate things, marital violence, partner infidelity, teenage pregnancy right down to the death of beloved pets. I find it difficult to feign lack of interest, because I'm fascinated. These are real life soap operas. However, getting involved kills the "Employer/PA relationship" stone dead." (Vasey, op. cit.)

With a PA in attendance, there can be problems going to the cinema, or in an intimate dinner, as the competing needs for assistance and for privacy are balanced. There are other issues as well. One tetraplegic lost a friend who could not understand why a black person was doing all the menial work. A gay tetraplegic had a problem when his equally gay PA was flirting in a gay bar when he should have been working for him. There may also be ethical judgments; what if the tetraplegic wanted to smoke dope, should a PA role the spliff? If a person lacks independent action, and if society is happy for someone else to provide it, then how should we set limits to a person's independence of volition and action? Overall, however, the use of PAs has allowed far more than living outside a care home; it has enabled a return of agency, spontaneity, and choice, incalculable parts of self-esteem.

5.4 Albert Bull

Despite all these factors, some with tetraplegia do poorly. The embodied self they are left with is insufficient and they cannot find a part of themselves to inhabit, feel good about and project into the world. But many – most do well and find ways of enjoying new things. What determines a person's outcome is difficult to quantify, though there are some pointers.

Stephen Bradshaw of the Spinal Injuries Association, and himself injured, is clear that one of the most precious gifts of all is good social skills. Tony, a C4 tetraplegic, and so without arm movement at all, was quite unpopular with the senior nurses on his ward, because the young nurses would hang round him rather than help the others. He had such an amazing social facility that, right from the start, he reached out to others, even during a long period on bed rest and in severe pain.

After six months in bed you need to be stimulated, so I would ask whoever was walking past for a chat. I would try and direct questions at them not about me. I would ask, "How's the family?" to promote talk.

Such talk is not facile and not simply an invitation to chat. It is saying that I, a tetraplegic, in bed and maybe in pain, am still human and still inquisitive about others, still a social being. Sir Ludwig Guttmann once said that the single remark from which he learnt most – and which he never forgot or ceased to quote – was by the Reverend Albert Bull, a paraplegic Army chaplain who had spent eighteen painful months at various hospitals before arriving at Stoke Mandeville in 1944. He wrote: "The first duty of a paraplegic is to cheer up his visitors." Guttmann felt that

This significant remark taught me and my staff how to educate the public to abandon its attitude of pity and replace it by positive and practical help in returning their paralysed fellow men and women to society. (Goodman 1986)

Those with spinal cord injury, and other neurological impairments, become defined in the eyes of people they meet by that impairment. They are always "that guy in the chair", never able to pass as being like the rest of us. Having to meet this and reach beyond it can be an enduring difficulty; not dealing with it but with us. One woman told me:

I used to go out on a Saturday night, but there was never any one interesting to talk with, just drunks I went to school with who would say how sorry they were. Please go away, I'd say, and please move that cigarette away from my shoulder. I don't like making an entrance and an exit. I hate to have a group of people lined up to say goodbye.

As Murphy (1987) suggests:

The totality of the impact of serious physical impairment on conscious thought ... gives disability a far stronger purchase on one's sense of who and what he is than do any social role, which can be manipulated. Each social role can be adjusted to the audience, each role played before a separate audience,

allowing us to lead multiple lives. One cannot however shelve a disability or hide it ... It is not a role: it is an identity ... society will not let him forget it.

The chair may not simply prevent anonymity, even amongst friends, it may also preclude a natural flow of conversation and social interaction. Sometimes with friends we might be enthusiastic and upbeat about something, and wish to make an entrance, at other times we might wish for a quieter time. The chair imposes on both by requiring the same inconveniences each and every time, flattening – or at least making more difficult – the projections of altered mood and some of the ease of conversation. Using Merleau-Ponty's phrase, it imposes a style. Much of the success of those with spinal cord injury may be viewed as a re-introducing of their own styles, with and beyond their injuries.

6. Conclusion: Doing into Being?

For Merleau-Ponty, one of the philosophers most interested in the relation between agency, action, embodiment, and the self,

Consciousness projects itself into the physical world and has a body ... [It] is in the first place not a matter of "I think that" but of "I can." Consciousness is being-towards-the-thing through the intermediary of the body...

... my love, hatred and will are not certain as mere thoughts about loving, hating and willing: on the contrary the whole certainty of these thoughts is owed to that of the acts of love, hatred and will of which I am quite sure because I perform them ... I make my reality and find myself only in the act ... It is not because I think I am, that I am. The whole certainty of love, hatred or will is that *I perform them*. [My italics.] (Merleau-Ponty 1962)

It is true that many with tetraplegia are unable to act on the world, with profound consequences for their sense of being, for their self-consciousness, and their self-other balance. But they reveal something of the creative imagination in managing as best they can and in the return of a perception of normal to their situation. Samuel Beckett (1957) wrote that "You do what you are, you do a fraction of what you are, you suffer a dreary ooze of your being into doing." This dreary ooze is a problem that those with spinal injury might be expected to struggle with each day. Yet David, a tetraplegic, suggested that the playwright was mistaken.

I think he got it the wrong way round – Yes there are ways of being other than just doing, and probably the more creative you are the better. For me, though, it helps to explain my self-perception and self-image by saying that

the key is doing into being – the opposite of Beckett. Dragging more out of my head into my body, spreading my physicality throughout my body, gives me the possibility of physically doing things. Being into Doing – I don't think that's right. I think the things that I do make me who I am. That need to express myself physically in certain things leads me from the doing of things into being a person.

Merleau-Ponty (1962) suggested that “freedom is doing ... Once my actions cease to be mine, I shall never recover them, and if I lose my hold on the world it will never be restored to me.” In this, in relation to spinal cord injury – fortunately – he is now wrong. PAs return agency and action to their employers, so enabling independence and increasing spontaneity and freedom, and choice in whether to work or not.

Tetraplegics know the necessity of doing into being and the importance of action and embodied existence for self-esteem and self-respect. In their coming to terms with such an extreme dislocation in the “absent body”, they reveal much about the importance of the embodied self, its relation with the world and as an intermediary for social relatedness with others. In the ways in which the impairments consequent upon SCI can be accommodated lies an expression of the creativity and imagination within us, and which most of us – not so tested by illness and loss – thankfully, may not have to draw upon. They have to come to terms with a dramatic and sudden change in their bodies, care for a part of themselves they can neither feel nor move, learn to navigate their way through the physical world and to maintain and develop their personal relationships. Each and all of these involve dislocations in the normal flow of life and reveal something of how self-esteem, self-regard, and a consciousness of self, in these terms, is dependent on the body.

Notes

1. Tetraplegia is the term used in the UK. In the US the term “quadriplegia” is employed, though this combines Greek and Latin roots. When discussing spinal cord injury with people, I originally always used the politically correct “person with tetraplegia”, thus defining the person first and independent of his/her impairment. In fact, some with the impairment prefer to be defined as tetraplegics, proud of their identity as such. So I have used the terms interchangeably.
2. Since men outnumber women with spinal cord injury by approximately nine to one, I hope it will be acceptable to use the male pronoun as a default.
3. Men with complete tetraplegia may experience spontaneous erections, through intact spinal reflexes in the sacral cord, but may be unable to become erect in the usual fashion.

4. Before the emergence of long-term care for those with spinal cord injury, these people would almost invariably die of infection and, literally, used to rot away on back wards of hospitals. Even today pressure sores are a not uncommon event in tetraplegia and sometimes require months of lying on a bed resting the area.

References

- Beckett, Samuel (1957). *"Murphy"*. New York: Grove Press.
- Boks, P. (2003). *Into the silent land*. London: Atlantic Books.
- Brissenden, S. quoted in Oliver (1990). *The politics of disablement* (p. 173). Macmillan: London.
- Cole, J. (1995). *About face*. Cambridge, MA: The MIT Press.
- Cole, J. (2004). *Still lives; narratives of spinal cord injury*. Cambridge, MA: The MIT Press.
- Damasio, A. (2000). *The feeling of what happens*. London: Heinemann.
- Finkelstein, V. (1988). "Changes in thinking about disability", unpublished manuscript (4–5). Quoted in Oliver, M. (1990), *The politics of disablement* (Chapter 4). London: Macmillan.
- Gallagher, S. (1986). Lived body and environment. *Research in Phenomenology*, 16, 139–170.
- Goldstein, K. (1939). *The organism: A holistic approach to biology*. Boston: Beacon Press.
- Goodman, S. (1986). *Spirit of Stoke Mandeville: The story of Sir Ludwig Guttmann*. London: Collins.
- Hill, J. (2000). *Footprints in the snow*. London: Macmillan.
- Leder, D. (1990). *The absent body*. Chicago: University of Chicago Press.
- Merleau-Ponty, M. (1962). *Phenomenology of perception*. London, Routledge, New York: The Humanities Press.
- Morris, J. (1991). *Pride against prejudice*. London: The Women's Press.
- Murphy, R. F. (1987). *The body silent*. New York: Holt.
- Oliver, M. (1990). *The politics of disablement*. Macmillan: London.
- Royal College of Physicians of London and The Prince of Wales' Advisory Group on Disability (1992). *A Charter for Disabled People using Hospitals*. London: The Royal College of Physicians.
- Sacks, O. (1985). *The man who mistook his wife for a hat*. London: Duckworth.
- Shoemaker, S. (1976). Embodiment and behavior. In A. Rorty (Ed.), *The identities of persons* (pp. 109–137). Berkeley: University of California Press.
- Stover, S. L., J. A. DeLisa, & G. G. Whiteneck (1995). *Spinal cord injury. Clinical outcomes from the model systems*. Gaithersburg, MA: Aspen Publishers.
- Weiskrantz, L. (1997). *Consciousness lost and found: A neuropsychological explanation*. Oxford: Oxford University Press.
- Varela, F. J., et al. (1991). *The embodied mind*. Cambridge, MA: The MIT Press.
- Vasey, S. (1999). *The rough guide to managing personal assistants*. London: National Council for Independent Living.
- Zola, I. (1982). Social and cultural disincentives to independent living. *Archives of Physical Medicine and Rehabilitation*, 63, 84.

CHAPTER 7

Self and identity

Arne Grøn

University of Copenhagen

1. Concepts of identity

Imagine that I am in a hurry to catch a train departing at 7:52. It is early in the morning and it is rush hour, but I find my way to Copenhagen Central Station. I look at one of the screens that displays from which platform the train departs. I hasten to platform 5 where I find the train. I look at the car numbers, enter number 14, and finally I find seat number 62, where, according to the reservation ticket, I am supposed to sit during this journey.

As this sequence of situations, events, observations, and acts indicates, we orient ourselves in terms of identity. In the situations described, identity is a matter of identifying something particular by picking it out, be it the Central Station, the train departing at 7:52 from platform 5, car 14, or seat number 62. When I find what I am looking for, I could say: “*This* is the train” or “*This* is the seat.” If I visit the city for the first time, it might be a problem to identify the Central Station, or maybe even to identify this railway station as the station from where the train departs (as, e.g., in Lille, where you might go to Lille Flandres just to discover that your train departs from Lille Europe). We identify *this* particular building or *this* particular train by distinguishing it from other buildings or trains. Thus, implied in my search for orientation is the identity question: “What is this?”

It is easy to imagine that the train is delayed, so that when it arrives, I might ask: “Is this the train that was scheduled to depart at 7:52?” Maybe it even departs from another platform when it is delayed. In order to know if it is the *same* train, my question would have to make the destination explicit: “Is this the train for Hamburg that was supposed to depart at 7:52?” So, to push the sequence a bit further, as a mini narrative, imagine that I am heading for Berlin

via Hamburg, and that I have not been in Berlin after 1989. If upon my arrival in Berlin, I now walk to Potsdamer Platz, I might ask myself: Is this really the same city? Here the issue of identity is no longer simply a matter of identifying something by picking it out, distinguishing it from other objects, as in our first instance of the identity question, but a matter of re-identifying or re-cognizing this city as the same city I visited many years ago. To change the story a bit, imagine that I had visited Berlin in the meantime, for example in 1997, and at that time seen that the whole area around Potsdamer Platz was turned into a huge reconstruction site. From that visit I'll carry with me a question as to how this area will come to look in the future. This would now help me to realize that this part of the city has changed almost beyond recognition. Also, as I now walk from Potsdamer Strasse to Potsdamer Platz, I have some points of reference. If such points were missing, the situation itself would change. Either I would not be able to recognize what I now see again, and I would feel lost, or, knowing where I am, I would recognize it with an intense sense of change. To give an example that turns the direction of recognition backwards, a couple of years ago I saw some photos showing streets and places in Rotterdam before 1940. There were pictures of streets that I thought I knew, but which I could not recognize. There were no points of reference that made it possible to see what had changed. As other photos of the inner city being almost completely destroyed in 1940 showed, this was due to the history of the city itself.

In the second instance, the identity question takes the form: "Is this the same as..?" Thus, we can distinguish between two concepts of identity. Corresponding to the *first* instance, we can speak of synchronic identity (for example, I face two options: Is the railway station where my train departs Lille Flandres or Lille Europe?). The issue is here to identify something by picking it out as *this* particular object, e.g., as this particular building, in contrast to other buildings or to other possible identifications of this building that imply that it is not the building I am looking for (e.g., the building turns out not to be the station from where my train departs, or it turns out not to be a railway station at all). In the *second* instance, the identity in question is diachronic. It is identity *over time*.¹ The issue here is to identify something particular as the *same*: Is this (really) the same city I remember having visited some years ago?

In both cases, the question of identity is implicit in our understanding of the situation in which we seek to orient ourselves.² The issue of identity is thus fundamental, but it is also puzzling. Identity implies difference. In order to identify something particular, we must distinguish it from something else (other buildings, cars, seats, etc.). And to identify something as this particular object not only implies that it is different from something else, but also that

it itself can be described differently from what I now identify it as. I might be mistaken; the building might turn out to be something else than I anticipated. Or I might be looking for something else; for example, I might see the building as a piece of architecture. However, these various ways of describing the same object, e.g. the Central Station, need not complicate each other. It is the same object in question. Thus, in the *first* instance where we identify something as *this* particular object by picking it out, the issue of identity might not seem that puzzling. In the *second* instance, however, the relation between identity and difference turns out to be more intrinsic and thus problematic. Here, identity is not a matter of identifying something particular by picking it out, but of re-identifying it as the *same*, although it has changed. This might give rise to the question what it is *in itself*. In this move, the *problem* of identity is accentuated. If identity is what a thing is despite its change in time, how do we account for the fact that it is this particular thing that has changed? This way of accentuating the problem of identity seems pertinent when the particular something in question is itself a *self*. If it is a self, it should make sense to ask: Is it the same as *itself* when it has changed?

Thus, we have finally reached the theme of this article. As the long prelude suggests, the title could have been different. In fact, it was first announced as a paper on “Concepts of Identity”, but as you can see, the title is no longer the same. The article will deal with self-identity, as the identity of a self, but it will do so in focusing on the very approach to the issue of self-identity. The main argument will be that the problem of self-identity is part of being a self. The difficulty in approaching the problem of self-identity tells something about what it is to be a self. Therefore, the title is “Self and Identity”, and not simply “Self-Identity”.

2. Two approaches

What is the problem of identity? It is often, without further reflection, taken as the problem of ascertaining or deciding the identity of something we observe, be that some thing, some event, or some other person. We then ask for criteria of identity. If what we are considering is the identity of a person, we may then discuss whether the specific criteria are mental (psychological) or physical. This discussion has as its *locus classicus* John Locke’s *An Essay Concerning Human Understanding*, book 2, Chapter 27,³ where the memory version of the psychological criterion of personal identity is put forward: personal identity “consists” in consciousness⁴ and reaches as far backwards as consciousness does.⁵ If we,

e.g., consult the entry on personal identity in the *Routledge Encyclopedia of Philosophy* (Garrett 2001), the issue of criteria of personal identity – whether they are physical, psychological or mixed – seems to be all there is to it. The problem of identity then is how to decide that two occurrences are instances of *the same*. In the case of a person, the question is, to quote the article on personal identity in *The Encyclopedia of Philosophy*:

What enables us to say, in spite of the changes wrought by time, that person A, before us now, is the person B whom we formerly knew and that person C, also before us now, is not? (Penulham 1967:95)⁶

The identity question then runs like this: Is what we are observing now the same as what we have observed earlier? In the case of personal identity, how do we establish that the person before us now is the same as the person we knew before? However, this is only a problem because something else makes it a problem. What comes in between, making it possible to speak of two instances of the same, is *time*. The problem of identity leads back to the problem of time changing our world, others, and ourselves. Identity then seems to be *sameness* over time or permanence in time. If this is what personal identity is, we can also look at ourselves in this way: How am I to decide that I am the same person now as I was one year or 10 years ago? What are the criteria for my identity as continuity in time?

The rest of the article could be a discussion of the problem of criteria for personal identity. But let us ask the question once more: What is the problem of personal identity? Is it the problem we deal with when we are looking for criteria of identity? Is this the way we encounter the problem of personal identity when it in fact is a problem for us?

I would suggest the following take on the problem. If we move in the opposite direction (not asking how to identify another person or ourselves as the same, but starting by asking what it is to be a self), the problem of self-identity turns out to be located at a more intrinsic level. It is a question of what it is like to be a self. The crucial point is that the problem of *identity* is part of being a *self*. We cannot account for what it is to be a self without taking into this account the problem of identity. What we would describe as a self is what it is *through* the problem of identity. This means that the problem of identity is not simply a matter of description (to be dissolved if we describe what we observe in an adequate manner). On the contrary, an adequate description would have to deal with the problem of identity. Self-identity is a matter of self-understanding.

Consequently, we can distinguish between *two approaches* to the problem of self-identity. According to the first approach, the problem is how to decide the identity of a self, asking for criteria of identity, be that physical, psychological, or mixed. According to the second, the problem is intrinsic to selfhood. The second approach is more complicated, taking into account that to be a self is already to identify oneself.

3. Identity and identification – and self-identification

In the prelude, identity was taken as a matter of identification or re-identification. To the question, “What is this?” corresponds “This is ...” and to identify “this” is to tell what the thing in question is. The question of identity was implicit in a process of identification through which we orient ourselves in a given situation. If we cannot orient ourselves, the question may become explicit. The situation might then become “reflected” in the sense that we ask what it is about.

Let us now turn to the issue of self, identity, and identification. If what we orient ourselves in relation to is an other self, the situation itself can be at issue. The situation can be shared, but also questioned, by the other. This means that the relation between identity and identification is complicated. Is the identity of another person a matter of identification? Normally, we do not check the identity of others, unless it is our job to do so. If I am going to meet someone at the station where my train arrives and if I have not met her before, I’ll first have to identify her. In this sense, her identity would be a matter of identification: I identify her by picking her out, distinguishing her from others at the station. But if I am going to meet someone whom I met several years ago, I do not identify or re-identify her, I *recognize* her. As we have not met for years, this might be difficult. The difficulty, however, is most likely not a matter of identification. Instead, the situation is marked by a peculiar ambiguity: I recognize her *and* see that she has changed. This situation is probably shared by the other so that the same takes place on the other side. We might both say: “It is a long time ago,” but we might also add, “You have not changed. You are still the same ...”.

The question of identity thus changes when we ask whether another person is the same. In most cases it is not a matter of simply re-identifying the other as the same person we knew before, but of *understanding* her. If the other has changed almost beyond recognition, the difficulty consists in understanding that she is still the same person we can identify her as. We know her to be the

same, but the problem is precisely to understand this – *that* she is the same and *how* she is the same. This is especially the case if she has changed in ways that we sense to be a change of character. The situation would then be marked by a more intense ambiguity. If she feels that we consider her to have changed beyond recognition, she might herself take this as a matter of identity in terms of self-understanding. Her sense of identity might be affected. To tell someone that she is no longer the same person she used to be would change the situation. As already indicated, the other can reverse the situation. She might question our way of seeing her, and even our self-understanding. We might then meet in a sort of reflective shared understanding. We know each other to be the same person we knew before, but for both of us it is difficult to understand how the other is the same.

Recognition of others takes place in a context of orientation that is affected by our relation to others. The question is, then, what we are looking for when we ask whether something is the same. Take again the example of a city, or part of a city, that has changed dramatically. What is meant by asking whether this city, or part of the city, is the same? It can be a matter of simple re-identification, but this would probably only give the point of departure. The question would then be: Would it be quite another city to live in? Has the atmosphere changed? We might even say that the city has changed its character.

What about one's own identity? What is meant if I ask whether I am the same as I remember myself to be 10 years ago? Normally, it does not make sense to apply criteria, for example the criterion of psychological continuity in terms of memory, in order to identify or re-identify oneself. We do not take ourselves to be a particular object to be identified. The very act of applying criteria, and indeed the very act of remembering, presupposes what the criteria should establish. One could argue that in cases where a person suffers from a loss of memory, the sense of identity might be questioned by the person herself. But still, this act of questioning presupposes some minimal sense of identity. If this is the case, to remember would not be to re-identify, but to recognize or to appropriate oneself in the past. In this, however, the ambiguity can be a part of the sense of identity. One senses oneself *as* changed so that one is left bewildered (in not being able to remember). This means that a minimal sense of identity is not some sort of basis for recognizing oneself as the same. It is more a sense of what is at stake in recognizing oneself. A minimal sense of identity is already implied in recognizing oneself as the one that has difficulties in remembering, to the point of bewilderment. If one cannot remember past experiences one is supposed to remember, by others or even by oneself, and

if this affects one's way of living, this may even become part of one's sense of identity ("I cannot remember").

The sense of identity can also be ambiguous when one actually remembers past experiences and past acts as one's own. I can identify myself as the one who 10 years ago wrote the letter that I now read, and still I can have difficulties in understanding *that* it is me. I recognize that I am the same as the one who wrote this letter, but this is exactly what is difficult to understand. To recognize, then, is a more complex undertaking than to identify (in the sense of picking this particular one out) or to re-identify (in the sense of establishing this to be the same as): I can recognize in the sense of identifying myself as the same, *and* have difficulties in recognizing this, in the sense of understanding and maybe accepting this. One might even say that it is *as if* it was someone else who wrote the letter. The answer to this bewilderment is not a re-identification. On the contrary, the fact that I am the same, in the sense that I wrote the letter, is the point of departure in order to see the problem (if it is a problem – one might also be happy that continuity is in this sense lacking). What is required is a process of self-understanding or coming to terms with oneself.

What is the sum of these preliminary reflections on self, identity, and identification? *First*, when we ask the identity question ("What is this?" and "Is this the same as?") in relation to a self, other selves, or ourselves, the question itself is complicated. Identity is no longer a matter of simple identification or re-identification, but of recognition. If, however, we follow the first approach and look for criteria of identity, the identity question is taken as a matter of re-identification: How do we establish that this person is the same as the one we knew? What we then fail to understand is what it means for a person to change over time. If we are looking for personal identity over time, the issue is how we *as persons* change with time. We change through what we experience and do, and through how we, in this, understand the world, others, and ourselves. One basic feature is that we change over time through *relating to* time.

This leads to the *second*, and crucial, point: What makes the relation between identity and identification complicated, in the case of the identity of a self, is the fact that the "object" to be identified *already identifies itself*. If in the process of orienting myself the situation becomes "reflected", if my way of orienting myself is challenged, I encounter myself, not as yet another object, but as the one identifying, or as having already understood myself in trying to orient myself in the given situation. In relating to another person, I encounter her as someone who already identifies herself. This need not be in an explicit form as, for example, when she introduces herself to me. I can only understand her if I see that she already takes herself to be this person. This is part of what

it is to be a *self*. She is identifying *herself*, maybe in ways in which she directly distinguishes herself from others. This also indicates that she is able to reverse the situation: maybe she is trying to figure out who I am.

Consequently, the approach to the issue of self-identity, as the identity of a self, should reflect the fact that we as selves are already identifying *ourselves*. At issue is what it is to be a self. A human being is what she is in identifying herself. It might be argued that *identification* is not the right word. In a sense, this is what I have been arguing for. I do not identify myself as I identify or re-identify an object. But the reason for this is that *self-identification* is *radical*: taking myself to be this person implies that I am also the one to be what I take myself to be. In this case, identification is noncriterial. As a self, a human being individuates herself in thus taking herself to be what she is. This self-relation in self-identification is implicit when she distinguishes herself from others – or from their conceptions of who she is. The implication, however, is that self-identification can be problematic in ways that identification of others cannot. To identify oneself *as* has the ultimate implication to be accountable for (I should account for myself in ways I cannot, and should not, account for others). Furthermore, identification here can mean to identify oneself *with*. As selves, we can identify ourselves with and distinguish ourselves from others. Again, this need not take place in any direct manner; it might be implied in what we are doing.

Thus, the approach to the issue of self-identity should focus on this critical point: to be a self implies self-identification. Identity is not something given, which one then relates to in a reflective, higher order mode. We are not what we are and then, in a second move, understand what we are. Understanding is part of being human,⁷ and it is so in the sense that we are what we are in understanding – and misunderstanding – ourselves. Self-identity is a matter of self-understanding.

These reflections on self, identity, and identification are arguments for the second approach to the problem of self-identity. Self-identity is a problem for a self, and the problem is part of what it is to be a self. Thus, at issue is what it is to be a self. Let me propose the following definition: Self is to relate to oneself, and in relating to itself a self *is* self-relation.

4. Self-relation (1)

Self as self-relation is, I will argue, what is implied in the famous definition of the self in the opening section of Kierkegaard's *The Sickness unto Death*:

A human being is spirit. But what is spirit? Spirit is the self. But what is the self? The self is a relation that relates itself to itself or is the relation's relating itself to itself in the relation; the self is not the relation but is the relation's relating itself to itself. (Kierkegaard 1849/1980: 13)

Self is to relate to oneself. This definition is remarkable in at least two respects. First, it amounts to a de-substantialization. Self is not a substance, but a relation, or, to be more precise, self is a process: self-relating. Second, self is to relate to oneself *in* relating to others and to a more or less shared world. This relation is often ignored in interpretations of Kierkegaard. If self is to relate oneself, then a self is always a self *before* something, namely that which we understand ourselves in relation to. This can be taken from a less famous key passage in the second part of *The Sickness unto Death*: "The criterion for the self is always: that directly before which it is a self" (Kierkegaard 1849/1980: 79). Self is self in relation to that before which it understands itself. The criterion is thus located at the level of self-relation. The two points, taken from the two quotations respectively, go together: self is to relate to oneself, and self is self in relation to. One relates to oneself *in* relating to others and to a world in between.

In view of the problems of alterity and temporality, to be dealt with later, I would like to push the argument a bit further. One is only able to relate to oneself as another, and it is only possible to be oneself *in becoming* oneself. Both points underline the de-substantialization indicated in the definition. But if one is to become *oneself*, then becoming oneself implies relating to what one already is, in the mode of acknowledging: accounting for, taking responsibility for, be in charge of. And this presupposes that one *already* has oneself as a problem. In relating to oneself, one is already related to oneself. This is self as self-relation.

If the second approach to the problem of personal identity is to take identity as a problem inherent in being a self, and if self is to be understood as self-relation, what are the implications of this approach? Let me unfold some of the implications by re-formulating the difference between the first and the second approach, and let me do so by discussing Locke's grounding of the first approach. One might object that the first approach, where we look for criteria for personal identity, also concerns what it is to be a self. To decide whether someone is the same as we remember her to be is itself a matter of how this other person relates to herself and to others, including us. This, however, would be an argument in support of the second approach, to the effect that the problem of personal identity is to be located at the level where the other person relates to herself. To clarify this issue, let us have a closer look

at Locke's approach. The objection just mentioned may be strengthened by showing that, in Locke, the criterion of personal identity concerns the issue whether the other *herself* remembers past actions with the same self-awareness as she is now herself to herself. Thus, personal identity here seems to amount to *self*-identity.

It seems that Locke starts in the same way as the second approach in that he considers "what *Person* stands for". He answers that it "is a thinking intelligent Being, that has reason and reflection, and can consider it self as it self, the same thinking thing in different times and places" (Locke 1700/1979: 9). The question implied in this is how it is "*self* to it *self*" (1700/1979: 10). Furthermore, Locke's criterion for personal identity could also be interpreted as a way of de-substantializing the self: "This may shew us wherein *personal Identity* consists, not in the Identity of Substance, but, as I have said, in the Identity of *consciousness*" (Locke 1700/1979: 19).⁸ This could then be taken as follows: What a self is, is to be understood in terms of what it *does*, e.g., remembers. So where does the second approach differ from the first? It is only possible here to indicate some of the points at issue.

First, if we take Locke's memory version as a theory of what a self is, it is not clear what its claim is. If we extract that a person can "consider itself as itself", then Locke explains that it does only so "by that consciousness, which is inseparable from thinking", and "by this every one is to himself, that which he calls *self*." Consciousness is that which "makes every one to be, what he calls *self*" (Locke 1700/1979: 9). Self or person then seems to be what comes out of this consciousness, or consciousness is supposed to explain what a person or a self is. But consciousness itself "consists in" considering oneself as oneself, thereby distinguishing oneself from all other thinking things. In other words, consciousness is the awareness and the memory *to be oneself*. This is not only an argument to the effect that consciousness of personal identity "presupposes, and therefore cannot constitute, personal identity" (as Bishop Butler argues),⁹ or that Locke's memory version of the criterion for personal identity is circular. The argument is that it does not account for what self and memory are.

Let us take memory as the *second* point. According to Locke, to be a person is to consider myself as myself, and I do so only by that consciousness that "can be extended backwards to" (Locke 1700/1979: 9) past actions or thoughts, thus remembering myself in the past to be myself. However, this means that memory does in itself not yield identity. Either I remember that I in the past had the same consciousness as I now have writing this,¹⁰ and in that case, the criterion for my personal identity is, strictly speaking, my self-awareness now (e.g., that I now write this), which then can be extended backwards.¹¹ Or memory *is* the

awareness of myself being the same now as yesterday or 10 years ago, and in this case, the criterion is memory as self-awareness.¹² In both cases, identity is not established but presupposed. I remember *myself* doing and thinking, or rather having done and thought, that which I now remember.¹³ It could be argued that this is what Locke is aiming at. The issue is whether I can “appropriate” my past actions and thoughts (cf. Locke 1700/1979: 16). However, to decide my identity using the criterion of memory seems, according to Locke, to be a simple matter. What I cannot “by my consciousness make my own Thought and Action” will not belong to me (Locke 1700/1979: 24). But if memory is understood in terms of appropriation, the situation turns out to be more complicated.

Locke’s memory version of the criterion for personal identity does not account for the role that memory plays for a human being *as* a person or a self. In remembering, we relate to ourselves. We do not just observe what we remember. What we remember might matter to us. This is especially the case if what we remember is something we have thought or done ourselves. Remembering is something we do *as* selves. This is not only to be seen from the *ways* in which we remember, but also from the possible ways of *not* remembering. The past might have an unacknowledged presence, which can turn the present into the presence of this past. Therefore, the question is: What do we do when we remember – or forget? The problem of *identity* is to be located at this level, where we, for example in remembering, relate to ourselves in relating to others and to a world between us. Locke’s account of memory oscillates between simple re-identification and ethical appropriation. Or, rather, he does not see the difference, which means that he cannot account for what appropriation is. When I appropriate past actions and thoughts I relate to what I remember. I realize that I am the one having done and thought so. When it is established that I am the one having written this letter 10 years ago, the issue of identity is not settled. On the contrary, the *problem* of identity only begins at this point. It might be hard for me to realize what I have done. And if I cannot remember in the sense of appropriating my past actions and thoughts, this does not turn me into another person (as Locke would have it). On the contrary, it is a problem *for* me being the person I am. What is at issue in remembering then is to recognize, to acknowledge, or to appropriate what I myself remember to have done, thought, or felt in the past. The problem is precisely to understand that I am the same as I remember myself to be. Memory is not simply a way of establishing that I am the same as yesterday or 10 years ago. It is more a matter of coming to terms with oneself. In this sense, to remember is part of what it is to be a self. It is not simply to observe one’s identity, but a way of being oneself.

Third, we should then ask if what we find in Locke is really an account of what it is to be a *self*. At first it seems so. Locke's question is "what makes the same *Person*" or what "makes a Man be himself to himself" (Locke 1700/1979:10). Self, he further states, is "a Forensick Term appropriating Actions and their Merit" (Locke 1700/1979:26). Considering past actions and thoughts the question then is whether we are "concerned" and "accountable" (cf. Locke 1700/1979:16ff.). The criterion for imputing or attributing past actions to oneself is consciousness or memory. Only if we remember past actions as our own can we *become* "concerned and accountable" (Locke 1700/1979:26). But "whatever past Actions it [an intelligent being or agent] cannot reconcile or appropriate to that present *self* by consciousness, it can be no more concerned in, than if they had never been done" (Locke 1700/1979:26). At first, Locke appears to hold that to be a person or a self is to be accountable. On a closer look, however, what he says is that I am only accountable, and thereby a person, if I can appropriate the past actions I consider as my own. But what if I have lost the memory of some parts of my life? Am I not still the same person that did those actions I have now forgotten? Locke answers that I am only the same *man*, not the same *person*. This distinction allows him to engage in counting more selves or persons, depending on the (lack of) continuity of consciousness. If I do not remember what I have done because I was sleeping or because I was drunk, I am, on this criterion, not the same person. But why then is the same man punished in both cases? This is, Locke answers, because, in these cases, it is not possible to "distinguish certainly what is real, what counterfeit" (Locke 1700/1979:22). The same argument, however, would apply to the very criterion of memory. It is not possible to distinguish certainly whether it is real or counterfeit when someone claims not to remember.¹⁴

Again, Locke's account reduces the complexity inherent in being a self, at least in three respects:

1. If the same human being, on the memory criterion of personal identity, can be said to be different persons at different times, why does Locke not simply count these different persons as he would count other persons? Are they only numerically the same human being? Or does it actually matter that they are one and the same: *oneself*? To whom, we could ask, would it matter? To the human being who still is both persons: the one waking and the one sleeping, the one drunk and the one sober. It is, we could argue, one and the same life both persons relate to. The one sober is affected by, and maybe suffers from, what the one drunk does, and maybe the one drunk also suffers from what he himself does.¹⁵ In this case, we might say

that the same man is the same *person*: being concerned for himself. In contrast, if it is possible for “the same Man to have distinct incommunicable consciousness at different times,” this same man “would at different times make different Persons” (Locke 1700/1979:20). Would we say that these different persons are only numerically the same, or is the point that our notion of personal identity is put to test before the case of apparent split of incommunicable consciousness? Would we say that it is as if the same human being were different persons?

2. Locke’s distinction between man and person implies a distinction between person and body. Man is the living body, while a person can be vitally united to a body, but is in principle disembodied. The limbs of a self’s body are “a part of *himself*: He sympathizes and is concerned for them,” but if a hand is cut off, it is separated “from that consciousness he had of its Heat, Cold, and other Affections, and it is then no longer a part of that which is *himself*, any more than the remotest part of Matter.” Thus, Locke concludes, “we see the *Substance*, whereof *personal self* consisted at one time, may be varied at another, without the change of personal *Identity*” (Locke 1700/1979: 11).¹⁶ The self’s body can be substantially changed without the personal identity being changed. This however affects the very notion of personhood. How is it possible to remember past actions without being embodied? It is difficult to see in what sense a person is “concerned” for itself (cf. Locke 1700/1979:25) when personhood is reduced to self-awareness, and self-awareness to self-perception. For Locke, a person is not concerned for itself as a “man”, i.e., for its whole embodied human being. Locke’s account of selfhood oscillates between appropriation and indifference. In this sense, his notion of self is a “punctual self” (Taylor 1989:Chapter 9).
3. Locke’s own account presupposes a subject that is not accounted for. Who is it that “cannot reconcile or appropriate [past actions] to [the] present *self*” (Locke 1700/1979:26)? Is it not the same person as the one being accountable? Is it not the same person that can forget or pretend not to remember – and that is accountable for thus not remembering?

Let us then, as the *fourth* issue, consider Locke’s *approach* once more: (a) According to Locke, the criterion of the identity of a person is what “makes a Man be himself to himself” (Locke 1700/1979: 10). The criterion is the set of memories we have. But I do not use my memory as a criterion to decide whether I am myself or whether I am the same person now as I was 10 years ago (cf. Williams 1973:12ff.). Having memories, we are not in doubt about our identity in the

sense that we are the ones having these memories. It does not make sense to ask: How do I decide whether I am the same as the one I remember myself to be? If Locke's point is that I am myself when I remember myself to be so, this does not account for what makes me be myself to myself. If I ask myself whether I am the same person as I used to be, I am not looking for criteria according to which I can observe myself to be the same. The problem is to understand that I am the same that I can observe myself to be. The issue is to account for myself being the same. (b) If my consciousness of myself writing now is what makes me myself to myself, and if my memory (consciousness) now of having written this letter 10 years ago is what makes me the same person to me now, it is only so because, in writing now, I am already a self relating to others and to myself, and, having acted, thought, and felt in the past, I have already placed myself in relation to others and myself. Thus, when I remember I am already affected by my past. (c) Given this, let us consider the *situation* where we look for criteria of identity. How does Locke describe this? When I observe something past, the question is whether I can "upon recollection join [it] with that present consciousness, whereby I am now my *self*" (Locke 1700/1979:24). The identity question, however, should not only be directed to that which I consider, but more radically to the "I" in the question ("whether I can ..."), i.e., to the self relating to itself in asking the question. This is, in Locke's examples, the subject presupposed, which is not accounted for. If I am to decide whether I am the same, this would presuppose that I am the one to decide. For Locke, the situation is one of observation and findings:

And thus, by this consciousness, he finds himself to be the *same self* which did such or such an Action some Years since, by which he comes to be happy or miserable now. (Locke 1700/1979:25)

But we do not happen to find ourselves in this way.¹⁷ We do not in this sense come to be happy or miserable. We do not happen to be ourselves, but are ourselves in relating to ourselves. What we find ourselves to *be*, we are *as* selves, *in* relating, understanding, remembering, etc. What we find would not be *ourselves* if we were not already related to ourselves. The situation is one of searching for something. What are we looking for when asking about identity? If it is the assurance that

... we may always be sure that we are the same persons, that is, the same accountable agents or beings, now which we were as far back as our remembrance reaches: or as far as a perfectly just and good God will cause it to reach ...,¹⁸

the situation would be awkward. For what would it mean that we are the same persons? That we relate to ourselves as the same persons. It is not something we happen to be. That we are the same persons is presupposed in the situation where we ask about our identity. We are already selves in *finding ourselves* to be the same persons. If, in contrast, self is made into an object of perception, it can be an object of self-control and self-remaking. In this, once more, a self unaccounted for is presupposed: the self that takes the self as an object of perception, control, and re-making.

In sum, what is reduced in Locke's approach to the issue of personal identity is self as self-relation. When the issue of personal identity comes up, I am already related to myself in what I have done, thought, and felt in relation to others and a world in between. Let me unfold some of the implications of self as self-relation in three steps: alterity, temporality, and normativity.

5. Selfhood and alterity

The basic move (the definition of self as self-relation) involves a reinterpretation of otherness or alterity. This is a way of reformulating the *problem* of identity. As indicated, it is primarily described as the problem of how a person remains the same despite the passage of time. But there is more to the problem of identity in terms of otherness. The problem is not only to understand that the other is the *same*, over time, but also to understand that she is an *other* than oneself. Identity is also the otherness of the other. Furthermore, the problem of identity is also to understand oneself as another. If personal identity is a matter of remaining the same person although one has changed, the experience of oneself as another over time (e.g., seeing photos from one's childhood when the world was different) is absorbed into identity as *sameness*. But to understand oneself as another is to understand identity differently. Thus, let us have a look at the issue of identity and otherness, and proceed in three steps: first *self as an other*, second *the other as a self*, and third *dialectics of recognition* showing the interrelation.

5.1 Self as an other

The issue of self and alterity is implied in the definition of self as self-relation: Self is to relate to itself, but a self only relates to itself in relating to others and to a shared world. A self is self before an other in relation to which it sees itself. The implication is that one can only relate to oneself as an other than this

other. This opens up the problem of identity. In order to see this, let us once more consider the issue of self, identity, and identification.

Being a self, one already identifies oneself. As noted, identification here changes its meaning. It suffices to distinguish among three forms. *First*, I identify myself *as*, not as an object, not in a reflective stance, but in taking myself to be myself (this is what I called radical self-identification). If I ask whether I am the same person now as the one who wrote this letter 10 years ago, I already relate to myself. I already take myself in certain ways in order to ask whether I am the same person. Thus, I have identified myself *as* the one now having the problem of appropriating my past action. *Second*, identifying oneself can also take a more indirect form, where I identify myself *with* something other than myself. In so far as I identify *myself* with this, I also identify myself *as*. This indicates that we identify ourselves in relation to others. When I relate to others, I am implicitly relating to myself. I place myself in relation to the others that I relate to. This is more obvious if the relation is a matter of values. To identify oneself with, can then mean taking oneself to belong to: a family, a group of friends, a religious or non-religious community, etc. Indirectly, I identify myself *as* (e.g., as a member of the community). *Third*, self-identification can take a more explicit form, in terms of appropriating or acknowledging what I have done or thought.

The critical point is not only that one identifies oneself in relating to others, but that the *way* one does so is already a matter of self-relation. One can place one's identity, in the sense of self-regard or self-esteem, in relation to an other in such a way that one needs the other to confirm it. If a person takes herself to be loved by the beloved, she might identify herself *as* the one loved by the beloved, so that to herself, she would not be herself if not loved (cf. Kierkegaard 1849/1980: 19f.). In a sense, she identifies herself *with* the other, but in identifying herself *as* the one seen in this way by the other.

If self-identity can be at stake in relating to others, we need to reformulate the issue of selfhood and otherness. Identity is not to be understood as sameness, but as a way of relating to the other than oneself and, in this relation, to oneself as an other than this other. In this sense, the problem of otherness is intrinsic to selfhood.

5.2 The other as a self

Let us now turn briefly to the issue of *the identity of the other self*. The problem is not only to understand that the other is the same despite our different experiences of her, but to understand that she is an other than we see her as.

What it means that the identity of a self is its *self*-identity, we learn not only from ourselves relating to others, but also from others relating to us. If we speak in terms of identification, we relate to an other who already identifies *herself*: takes herself as herself in relating to us. What is more, the problem of identity is part of our life together. If identifying is part of sociality, what then is meant by the identity of the other self? What is it to identify someone who already relates to herself: telling us who she is – and in doing so relates to us, telling us who we are?

The identity of the other is a matter not only of cognition, but of *re-cognition*, which means not only of re-identifying the other as the same (self-identical), but of recognizing the other as an other than oneself (self-identical would here imply that the other identifies herself in relation to others, including myself).

5.3 Dialectics of recognition

This can be spelled out in a theory of recognition that is also a theory of self-consciousness and self-identity. It is only possible here to give a very brief sketch. Taking the lead from Hegel's *Phenomenology of Spirit* (1807/1952: Chapter IV A, on dependence and independence of self-consciousness), we can describe the intrinsic relation between selfhood and otherness as follows. The identity of a self depends on the recognition received from the other self, but what is to be recognized is the autonomy of the other. This means that self-identity is self-relation in relation to others. "Selbständigkeit", independence or autonomy, implies "in sich stehen", but this is realized in a relation to others, which is a relation of mutual recognition.

Further, the structure of *re-cognition* in the sense of acknowledgement or appraisal (German "Anerkennung") is double. To recognize someone *as* is not simply to re-identify him (e.g., "Oh, you are Mr. Stanley"). It is to affirm what one sees: the other as an other self. If re-identification is to confirm that the other we see is the same person as we knew before, this would in principle not concern the other as another person or self. In contrast, re-cognition means that we already have seen the other as an other *self*, and that we realize and affirm what this implies: that she relates to herself, has her own life, can reverse the situation in which I relate to her, and that it matters to me what she, as the other I relate to, does etc. Thus, a theory of re-cognition seeks to capture this interrelation between self-relation and relation to others. On the one hand, it is only possible to be oneself, independent, in relation to others (to be a self

is to relate to others). On the other hand, the relation between ourselves, as a relation of mutual re-cognition, is to affirm that the other is independent.

This interdependence of self and the other is also to be seen in the negative. This is what the *dialectics* of recognition in Hegel's *Phenomenology of Spirit* is about. The one, the master, seeks to be recognized without recognizing the other, the slave. But to be recognized by someone whom you do not yourself recognize is not recognition. The negative dialectics thus indicates the logic of recognition: recognition of self and the other is only possible as a reciprocal relation (Hegel 1807/1952: 142f.).

In our context, the double structure of re-cognition is crucial. The structure is also to be seen in the negative dialectics. As Jean-Paul Sartre once noted, it is impossible to treat a human being "as a dog" if one has not first seen him as a human being (1957/1970: 313).¹⁹ To deprive the other of significance implies that we have already seen him as a self. The identity of the other is then both seen and denied.

If otherness is intrinsic to selfhood, how does this alter the issue of self, identity, and identification? In the dialectics of recognition, the two issues go together: first, to identify *oneself* with the implication that one distinguishes oneself from the other, and in this comes to see oneself as an *other* than the other, and, second, to see the *other* as an other *self*, that already identifies herself.²⁰

Re-cognition is not re-identification. What is at stake in recognition is the identity of the self – both the identity of oneself as another (in relation to the other) and the identity of the other self. To the double structure of recognition corresponds a double implication of identity: (1) Identity depends on recognition. Not only do we understand ourselves in relation to others (a self is self before an other, before whom it sees itself), our self-understanding is affected by our (lack of) understanding the other as an other. (2) In recognizing, but also negatively in depriving the other of significance, we already relate to the identity of the other self. In this sense, identity is not what comes out of, but what is affirmed in, re-cognition. We need to take both implications into account in order to understand what identity is. Self-identity is not something inside the self, but is at stake in the relation to others in that it is a matter of relating to the other. However, it is also a matter of the identity of the self *in* this relation. This double implication is to be developed in interpreting the normative dimension of self-identity. But before that, we need to consider the second part of the *problem* of self-identity: self as another in time.

Until now, in discussing selfhood and otherness, I have abstracted from the issue of temporality. However, in order to understand how the notion of iden-

tity is changed, we need to take the temporality of the self into account. One is another in relation to the other self, and one is another in relation to time.

6. Selfhood and temporality

Although the issue of identity, as we have seen, is not to be reduced to the problem of *time*, still the *problem* of identity is, in a critical sense, time. The problem of personal identity reflects the problem of the temporality of the self. What it means to experience time and to live temporally, is to be seen through the problem of identity. In the *first* approach, however, where we ask for criteria of identity as sameness, the question is identity over time as permanence in time (what remains the same in time). In contrast, the *second* approach focuses on temporality as an intrinsic feature of selfhood. We cannot explain what it is to be a *self* without taking the *problem* of time into account. As selves, we live through the problem of time. In this sense, we might ask ourselves whether we remain the same over time, but this question has then changed its meaning. It is, to put it concisely, not a matter of observation, but of orientation and obligation. We do not observe ourselves to be the same, despite the change of time. The change of time affects us more basically, in our self-relation. We are ourselves in relating to time and we relate to ourselves in time.

What are the implications for the notion of identity? Identity over time is not to be understood simply as sameness or permanence in time, despite of time, but, more intrinsically, as a matter of selfhood. If, as a self, a human being is only herself in becoming herself, then time is not just an obstacle to identity, something to be overcome, but a condition of identity. *Memory* is a way of appropriating ourselves in time changing us. In this sense, to remember is a way of being oneself across time. *Character* might then be reinterpreted as accumulated life history in the sense that it reflects how the individual has taken her history. Thus, character is not to remain the same in spite of time, but to deal with time's change. Time would still be a challenge to identity, or, rather, personal identity is to be understood *as* challenged by time. Time changes us, but we change in time *as* selves. This is not only to be seen where we (if ever) seem to succeed in making time into our own, but also where we fail to do so. Time can change us, and we can change in time, so that we lose ourselves in forms of despair. But still, such negative forms of temporality are ways in which we *as* selves can change. If one's life falls apart, it is still one's life, it is still a way to be a self. Even if I have changed in ways that are difficult for me

to appropriate or to reconcile (to use Locke's words), I am the one who has changed. To see this might be the critical point in order to change myself.

One could then say that identity is not something given, but a matter of becoming. I am only myself in becoming myself. This process of becoming, however, is itself a matter of self-relation: I am to become *myself*. But this opens up possibilities of *not* becoming oneself in the sense of appropriating oneself or reconciling with oneself. The implication is that identity is *at stake* in this process. This is part of what it is to be a self.

The shift in the notion of identity, using Paul Ricoeur's terminology (Ricoeur 1992), could be formulated as follows: Self-identity is not to be understood in terms of *sameness* (*idem*), but in terms of *selfhood* (*ipse*). This, however, is only a first step. It must be followed by a second, explicating the normative dimension of self and identity.

7. Normativity

The normative character of self-identity is crucial for understanding the issue of self and identity. However, it is not clear in what sense personal identity is normative, and what the implications are.

First, the normative character is not to be taken as an ideal or a standard, "the true self", but more basically as the normative dimension of selfhood in the sense that the identity of a self is at stake: it can be respected or violated, recognized or overlooked, affirmed or denied. What is respected or violated is not an ideal self, but the other as a self in the sense that she is *already* herself. But identity as self-identity is also something to be achieved. In this sense, it implies a standard, but this standard is only to be met, that is: identity to be achieved, if one appropriates oneself. I am to become *myself*. This means that identity also in this normative sense (as something to be achieved) is a matter of self-relation. The standard itself concerns *how* I seek to realize it in relating to myself.

Second, self-identity is at stake both in the case of recognition (in relation to the other) and in the case of appropriation (in relation to time). In the first case, identity is a matter of re-cognition, as we have seen. On the one hand, self-identity is affected by (lack of) recognition. This has to do with the nature of self-identity: that it concerns what it is to be a self being oneself. To be a self is to relate to oneself. This means that one only 'has' self-identity in self-understanding: identifying oneself in the sense of taking oneself as, and appropriating oneself. Thus, when one's identity is affirmed and recognized, one

can see oneself as a self recognized. On the other hand, identity is not simply what comes out of recognition. If it were, we would not understand what recognition is. Recognition is about identity. The normative dimension of identity comes to the fore in the demand of respecting and recognizing identity. What then are the implications for the notion of identity? In a sense, the identity of a self is something given: to be recognized, but it is not something given in the sense of a substance, an inner or true self. Instead, identity of a self is fragile and vulnerable. This has to do with what it is to be a self: a self already takes itself as a self. This implies that it can be affected as a self, in its self-understanding, self-regard, or self-esteem. Self-identity is fragile as a matter of self-understanding.

We are not just what we are, as some sort of substance, but neither is identity a matter of constructing identity. To our identity as selves belongs that we change, and that we do so as selves. Something is at stake for us in time changing us. In understanding oneself, one relates to the life that one has to live, in ways in which others cannot and should not. This is part of one's self-identity. In this sense, self-identity has to do with self-esteem and dignity. To respect the identity of the other self is to have a sense of dignity, not only the dignity of the other, but also one's own. Conversely, one's own dignity is not only a matter of how one conducts one's life, but also concerns how one relates to the lives of others.

This leads to the *third* point. I have touched upon the issue of self-identity and appropriation. One is to become oneself in the sense of appropriating oneself. This could be understood, in Locke's terms, as appropriating past actions and thoughts as one's own. In remembering them as my actions and thoughts, I am accountable. But self-accountability implies that I am both the one to account and the one to be accounted for. I should be one and the same: myself. This identity is a matter of concern or commitment. I do not observe myself to be accountable, I am to be accountable. And the past actions and thoughts I remember as mine are not isolated. They are part of a history that I have not just enacted or thought out, but have experienced or lived through. I am not only subject of past actions and thoughts, but subjected to my own history. I may identify this history as mine, and still be unable to appropriate it as my own. Furthermore, to take one's past upon oneself is a double move, it is backwards (remembering), but also forwards in the sense that it is to live a life with this past and these possibilities. This indicates that the *concern* for oneself implies *self-commitment*. Here the ethical significance of self-identification comes to the fore. We not only identify ourselves *with*, but commit ourselves *to*, so that the question will be whether we actually are, in the course of our life, commit-

ted *by* ourselves. To have a life to live as a self is a question of commitment. And to pose the identity question to oneself would here make sense, not as a matter of self-observation, but as a matter of self-obligation or self-commitment. This third point accentuates the normative dimension of self and identity, and the relation of normativity and *temporality*. Self-identity in time is also a question of committing oneself. This can be brought out in the case of keeping a promise. To keep a promise is to remain the same, not in the sense of permanence in time, despite of time, but through relating *to* time. The normative dimension of identity is the double possibility of fulfilment and betrayal.

Fourth, self-identification and identification between selves have, as already indicated, a profound, but also ambiguous, *ethical* significance. In order to bring this ambiguity out, we must discern a further layer of the issue of *identification*. In so far as self-identity takes place between selves, it is also a matter of how we see each other. This is the case in recognition. But recognition has negative counterparts. Thus, one can identify the other in the sense that one *has* decided who she is: she is what she *has* done. To judge or to condemn is in this sense to fix the identity of the other. The other is what one takes her to have shown herself as (e.g., she has shown herself to be a liar). This negative ethical significance of identifying, however, is to be set off against a positive one. To take an ethical stance implies seeing oneself as accountable for one's acts. The link between a person and her acts is in this sense crucial in ethics: I am to identify myself as the one who has done these acts. If I loosen this identity between person and act myself, I am in a sense not the same (I do not appropriate my acts), but then I am still the one who seeks to evade myself and in this sense tries not to be the same. Thus, the identity between a person and her acts is itself of an ambiguous ethical importance.

Fifth, the normative dimension is to be interpreted through *negative* phenomena. This is already implied in what has been said. The normative dimension concerns what it is to be what we already are: a self being oneself. It is a dimension for self-understanding, and not an ideal self set apart. It is a matter of identity in both directions: self-appropriation (I am to become myself) and recognition (of the identity of the other self). In both directions, we understand the normative implications of being a self in seeing the negative possibilities that we have ourselves. First, it is possible in various ways *not* to appropriate oneself. It is possible not to realize that I am the one accountable. It is also possible to lose oneself, in the sense that one's life falls apart and that one gives up oneself in despair (cf. Kierkegaard 1849/1980). Second, it is possible *not* to recognize the other in various ways. One can see the other in ways that deprive her of significance or dignity. But what self-appropriation and recogni-

tion mean should be a response to these negative possibilities. What is at stake in self-appropriation and recognition, we know from these possibilities that show the vulnerability of a self.

The crucial point here is that there is still self-identity in the negative instances. If I do not appropriate myself, I am still myself not appropriating myself. If I lose myself in despair, I am still myself losing myself in despair. We can say: "I am not myself," or "I am beside myself,"²¹ but still I am the one that is not myself. It concerns me.

Let us take once again the issue of remembering. What if I forget what I have done? According to Locke, it is not the person who forgets, for the person is accountable and is so only when she remembers. So the one forgetting is the man (cf. Locke 1700/1979: 20ff.). But to forget can be part of not acknowledging or not appropriating what one has done. Someone might say to me: "You should have remembered!" or "How could you forget?" In this, I am the one forgetting and might be accountable for forgetting. I am accountable *as* the one forgetting. If we should put it in Locke's terms, it would then be the man who is the self. Or, to be more precise, *self*-accountability presupposes that we can fail *as* selves, not only in the sense that we should account for our failures, but also in the sense that we can fail, maybe resist, to account for ourselves. Thus, in the demand to account for ourselves, self-identity is at stake in a critical sense. We are ourselves questioned as the ones to account for who we are.

This leads to the *sixth* point in the attempt to explicate the normative implications of self and identity. This last point can be put in terms of a brief summary. I have argued that the normative dimension is not separate, but concerns what we are as selves. The implication is *double*: the normative dimension is that the identity of a self is *at stake*, but this presupposes that it is *already* a self in the sense of self-relating. This can be seen in both directions (self-appropriation and recognition). What I should account for or appropriate is myself, which implies that I am already a self in the sense that I can fail. And what I should recognize is the other *as* an other self to be recognized. In recognizing, I relate to what I have already seen: that the other already is another self relating to herself. As mentioned, this can be seen negatively. To humiliate another person implies to have seen her as someone that can be humiliated, that is, as another person. Or, to use Sartre's example, to treat a human being as a dog implies that one has seen him or her as something else: as a human being.

8. Hermeneutics of the self

Let us now return to the approach to the issue of personal identity. The approach should reflect the nature of the problem. I have argued that the problem of self-identity is inherent to selfhood. The identity of a person is a matter of how she relates to herself: how she identifies herself, takes herself, understands herself. She can reverse the situation of identification or recognition, not only in identifying herself before us, but in relating to us as the others identifying her. She might question what we do in identifying her. Thus, what complicates the approach is the fact that the *problem* of identity is already part of being a person or a self. It is a problem for oneself, it is a problem between ourselves, and it is a problem about ourselves. The identity of a self is *self*-identity.

This approach to the problem of personal identity could be called hermeneutical. It emphasizes that self-identity is a matter of self-understanding. Self-identity is a problem for a self seeking to understand itself. Furthermore, a hermeneutical approach focuses on the self being situated. It is embodied in being embedded in a social, cultural, and historical context, and it is situated *as* a self, in understanding and relating to the situation. If we are selves in relating to our world, what are the implications for the notion of identity?

Personal identity is not simply to remain the same. To ask whether I am the same over time, according to Locke, can be taken as a matter of concern. The concern for my past identity harbours, so to speak, a concern for my future survival. My identity matters to me. But my identity is itself a matter of concern in the sense that it *depends on* what is valuable to me or what I am concerned about. We decide what matters to us, but we are also affected, or even defined, by what matters to us. Thus, there is a critical difference between saying that our *identity* is what matters to us and saying that our identity depends on *what* matters to us. In the latter case, identity is itself a matter of what matters to us.

It might take some time, maybe most of a life time, to find out what matters to us. It is part of our life to ask what it is about. When we try to find out what matters to us, our past is also a future resource. We can reflect on experiences from which we have come to understand what matters to us. In remembering, we are already affected by what has happened to us in the past. We carry with us a history without which we would not be what we are. This does not mean, however, that we are our history. We relate to it, also when we try to forget or leave our past behind us. We can carry, on the other hand, our history with us, without appropriating it as our own. Thus, we would not be ourselves without our history, which we also can leave behind. Self-appropriation takes place in relation to a history, but it can do so in complex ways. We can seek to remember

what was important to us; we can seek to appropriate our past in order to free future possibilities, and we can remember possibilities that were cut off in the past in order to change our future.

Thus, the issue of self-identity and history should not be restricted to the identification of oneself in one's past actions and thoughts. We tell (parts of) our history in order to tell who we are. We identify ourselves by our history: where we come from, what we have experienced and done in the course of our life, and where we plan to go. Self is a self on a journey. Identity is not something fixed, but to be developed. This is reflected in the stories we tell about ourselves. And if self-identity depends on what we identify ourselves as or identify ourselves with, one could argue that we not only tell, but also live, our lives as stories. This leads to the suggestion that in a hermeneutical approach, identity is *narrative* identity (cf. Zahavi 2003: 58). It could then be taken as the identity that is constructed through the narratives one can identify with. In Ricoeur, the notion of identity as *ipseity* is linked to the notion of narrative identity, but in distinguishing this from an ethical identity (Ricoeur 1992: 140ff.).

As mentioned, the approach to the issue of self and identity that I have outlined can be considered to be hermeneutical, but I would hesitate to call it narrative. Let me indicate just two arguments. First, our identity is not simply what comes out of the narratives we tell about ourselves. Identity is also how we relate to our narratives or to ourselves in telling these narratives. Second, the key notion of self as self-relation can be developed in a hermeneutical approach in terms of self and self-understanding. This approach would focus on how we implicitly understand ourselves in relating to others and to a world in between. In this sense, the identity of the self depends on how one understands oneself: how one takes oneself in taking what happens to oneself. Thus, how one relates bodily – in movements, gestures, seeing – to others is also to place oneself in a position to others. But this is not an identity that is developed in a narrative, but rather what narratives would relate to. In trying to understand what I experience by “constructing” some sort of narrative, I relate to myself as the one having experienced what I try to understand. Self-relation is not first to be established in a reflection or in a narrative, but is implicit in relating to others and a shared world.

9. Self-relation (2)

What then is the relation between a *normative* and a *minimal* notion of self and identity? In a normative sense, identity as self-identity can be appropriated or lost, respected or violated, as self-understanding and as integrity. In a minimal sense, self-identity is given in an awareness of my experiences as mine. This is a rather intricate issue to end up with, so let me just give two brief suggestions concerning the two notions that lead back to my main argument.

First, in what sense can we speak of a minimal self? If it is taken as a self not extended in time, in contrast to a narrative self (Gallagher 2000), in what sense is this minimal self *oneself*? What is the link between this minimal self and the one identifying this as oneself? My suggestion is that we can only speak of a minimal self in a non-minimal context. It might be an open question to me, what it means that I am the one who has this experience or does this act. This would in particular be the case if the experience or the act in question is of a more complex nature, e.g., making this journey in order to visit someone. The non-minimal context has to do with the implications of the experience or the act in question. These implications are not outside my self-awareness. In being aware of my experiences as mine, I implicitly understand the context and myself in the context. When asked, I can tell what I have done or experienced, and I do so by trying to make explicit what was implied in my experience or act, and in the situation. In a sense, I would try to tell what it means that this experience and this act are mine. That is, I would try to explicate my implicit understanding and my awareness of my experiences as mine. And I might do so in some sort of narrative. This could already indicate the link between a minimal and a narrative self.

Second, what would a non-minimal self look like? It could be taken as a narrative self, but, as I have argued, the core of the issue would be the problem of the normative dimension. I have further argued that a *normative* notion of self and identity implies a *non-normative*: to identify oneself as the one whose identity is in question or challenged. So, my second suggestion is that a non-minimal notion of self presupposes a minimal. The problem then is, once again, how to define this minimal self. The two suggestions can be rephrased as follows: A minimal notion of self is presupposed in a normative notion, which, on the other hand, offers a context for understanding the minimal self.

The *problem* of self-identity thus involves both a normative and a non-normative dimension of self and identity. If we distinguish between two levels, a minimal and a normative, the problem lies in between. On the one hand, a normative notion presupposes a non-normative: I am the one whose identity

is at stake. On the other hand, in order to make explicit what is implied in the awareness that a given experience or act is mine, we must ask what it is like to be the one whose experience or whose act it is.

Let me explicate the double suggestion – that a normative notion concerns the conditions or the context of a minimal, and that a normative presupposes a non-normative – by considering how a person with schizophrenia, or a schizophrenic person, can describe her/his self-awareness. This is not meant just to be an illustration. One of my aims in delineating a strong notion of self-identity that focuses on the normative dimension is to capture philosophically some of the challenges presented by psychopathology. The second approach to the issue of personal identity, and the corresponding notion of self-identity outlined here, should prove itself relevant not least in cases where the very conception of self-identity is challenged. Following the double suggestion, I'll move in both directions: from a minimal to a normative level, and from a normative to a presupposed self-relation.

In order to see how the two parts of this move are intertwined, let us take the argument about *normativity* and *negativity*, developed above, one step further. The point made there, namely that the significance of the normative dimension of self and identity is to be seen through negative phenomena, is accentuated when we are dealing not only with forms of self-relation that are *mis-relations* (as described in Kierkegaard's analyses of despair),²² but with *self-disturbances*. The acute question now is: What is the self-relation in *self-disturbances*? If self-disturbances only can happen to a self, in what sense are they ways of being a self, i.e., ways of relating to oneself? If we compare self-relation as mis-relation (as in despair) and self-disturbances, the difference seems to be the following. In self-relation as mis-relation (despair), I am still able to recognize myself as the one not being myself (for example by saying "I am not myself," "I am beside myself," or even, "I cannot recognize myself"). In self-disturbances, this is precisely what is at stake. The question here is whether one can identify oneself as the one being disturbed.

What is at issue here is what I have called radical self-identification: to see oneself as the one being in despair, or to identify oneself as the one suffering by or under self-disturbances, in distinguishing oneself from others. The structural disorders in schizophrenia seem to affect this minimal, but radical, self-identification. Let us take two examples of how a person with schizophrenia can describe her/his self-awareness:

- (1) "My 1st person perspective is lost, and replaced by 3rd person perspective;"

- (2) "It feels *as if* it was not me who is looking at the world from here".²³

How is the question of identity to be understood in this context? The *context* or the *situation* itself is complicated. In the second example, the one speaking relates to himself in the mode of *as if*, but this is a very complex way of relating to oneself. What is implied is a radical self-identification, which is at the same time at stake in what is said. One can ask whether, in the first example, the subject does not also speak in an as-if mode. To say that my first person perspective is lost, requires this first person perspective. If the interlocutor then asks: "It is as if your first person perspective were lost?", the situation might be changed in the sense that it would be a matter of recognizing oneself as the subject in assuming this as-if mode. On the other hand, if we take the second example, although it is *not* said that I am not the one who is looking at the world from here where I stand, yet the problem is to identify oneself *as* the one who is looking at the world from here. In both examples, the one speaking identifies herself (speaking of *her* perspective, or of *herself* as if not looking at the world from where *she* looks), but in a mode indicating that radical self-identification is a problem. Self-identification is being complicated. To identify oneself would be to recognize: *I* am the one who sees as if it was not me who was looking from here.

The question then is: What is the subjectivity involved or presupposed in this situation where a subject can talk of her perspective being lost and replaced by a third person perspective? The givenness of one's experiences as one's own is problematized, but in such a way that one is relating to oneself in this problematizing or as-if mode. To appropriate oneself would then be to appropriate oneself as the one looking from here where one looks. This means that self-appropriation would relate to what I have called the radical self-identification. This might seem paradoxical, because self-appropriation, as I have argued, would presuppose this self-identification. Here the hermeneutical character of self-appropriation comes to the fore: what is appropriated is presupposed or pre-understood. *What* is disturbed is the minimal self-awareness, one could still say, but it is disturbed *in* a context that is not minimal and that is implicitly understood. It could be objected that this is due to the fact that the situation itself is reflective, talking about how one is.²⁴ However, it takes more than a minimal self-awareness for this self-awareness to be a problem. If one's situation is in any way felt as a problem, the self-awareness is affected. In self-disturbances, the self being disturbed relates to itself in a complex, not minimal, sense. It encounters itself as being related to itself, as being burdened by itself. As the reflective situation (a form of speaker-hearer situation) indi-

cates, when a minimal self-awareness is turned into a problem, it is drawn into a normative context. What is at issue is whether I myself can see that I am the one, or actually see myself as the one, who sees in this as-if mode.

We have here moved primarily from a minimal to a complex level. When a minimal self-awareness is problematized, the issue is a complex one. Self-relation becomes a problem, which means that it is a problem to appropriate, acknowledge, or recognize oneself. The normative dimension of self-identity is then brought into focus, together with the context of the minimal self-awareness. But, so is the point made before: the normative dimension presupposes a non-normative notion of self. What is to be appropriated is oneself, in the cases just mentioned: the fact that I am already the one who sees in this as-if mode. The normative dimension thus pertains to the self that one already is in relating to others, to a world in between, and to oneself.

The notion of self *as self-relation* should capture precisely this relation between a normative and a non-normative dimension of the self. Self is not some sort of substance, but the fact that we relate to ourselves in being already related to ourselves. Self as self-relation is in this sense presupposed. Let me briefly substantiate this point by considering the critical approach to the notion of self in Hume and Nietzsche.

In the well-known section on personal identity in *Treatise on Human Nature*, Hume asked for “self or person” as “that to which our several impressions and ideas are supposed to have a reference”, and found:

For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception, and never can observe anything but the perception.

(Hume 1739/1896:251f.)

Self or person is “nothing but a bundle or collection of different perceptions” (252). What Hume is looking for is self as a substance. As he is not able to find a self as a substance unchanged in time or as a substance behind his perceptions, he reduces self to a bundle of different perceptions. But let us have a second look at the quote. Hume’s operation involves someone looking for “myself”. It thus presupposes a self that is not to be found where he is looking for it. If I take self to be nothing but a collection of different perceptions, I cannot account for the situation where I am looking for “myself”. What I am seeking to understand, “myself”, is neither a substance behind my perceptions, nor these perceptions, but the fact that I relate to myself, for example in looking for “myself”. The notion of self should account for this situation: that *we* are looking

for a self, in order to understand what we are, *as selves*. That we are selves is the implication of the situation. It is pre-understood in understanding the situation. What it means to be a self is thus to be seen from the very situation where we are looking for ourselves. The self presupposed is not simply an I-pole, but a self relating to itself.

Nietzsche not only criticizes the idea of self as a substance, he also holds the subject to be a fiction.²⁵ What is to be noted, however, is that *we* make this fiction (cf. e.g., Nietzsche 1988:465). Subjectivity then is to be looked for precisely in this self-relation: that we need the fiction of the subject in order to deal with ourselves.²⁶ This has to do with the problem of identity. A critical motive in Nietzsche is that we need the fiction of unity in order to cope with our own diversity. As Hume speaks of the self as a “commonwealth”, Nietzsche speaks of the subject as a “Vielheit” of subjects. The implication, however, is that we are already selves in having our own “Vielheit” or diversity as a problem. The subject is not simply a diversity of subjects. If we only consider a self to be diverse, we will have difficulties in accounting for the possibility that it can disintegrate. That one’s life is fragmented presupposes a minimal identity: that I am the one whose life is fragmented. If one cannot recognize oneself in this minimal sense, diversity will turn into disintegration.

In these two classic examples of a critical approach to the notion of self, self is presupposed as self-relation, in our understanding the situation in question. A *hermeneutical* approach to the issue of self and identity should explicate what is implied in this presupposition: self as self-relation. It can do so in focusing on the self-understanding that is already implied in being a self. What I have called the second approach to the issue of self and identity is in this sense hermeneutical. Self-identity is a matter of self-relation. The problem of self-identity is part of what it is to be a self. What we are is a matter of what we take ourselves to be: how we understand ourselves in relating to others and to a world in between.

In this sense, self-identity is a question to the self whose identity it is. It is a matter of self-understanding. Part of the solution to the quest for identity is that we are what we are in seeking to understand ourselves. If we could get some sort of definitive answer, this would change our situation as humans, our human condition. Thus, part of our human identity is the open question – about ourselves.²⁷

Notes

1. To the distinction between synchronic and diachronic identity, cf. Henrich (1979:140, 173).
2. The *problem* of identification comes up in situations in which we are not sure whether we are talking about the same thing (cf. Strawson 1959: Chapter 1). We can then try to identify what we are talking about. But identification also becomes a problem in situations where we lose our orientation. We might then become aware of the identifications we have already made in orienting ourselves. The context of orientation where identification becomes a problem need not be a speaker-hearer situation. But in the case of personal identity, the problem of identification is intrinsically related to situations of communication. The other self is also the self involved in a speaker-hearer situation in which we seek to identify what we are talking about. One of the aims in the following is to trace different meanings of identification.
3. As all references to Locke will be to this chapter, I'll only give the number of the paragraph referred to.
4. Cf. Locke (1700/1799:21): "... it is impossible to make personal Identity to consist in any thing but consciousness; or reach any farther than that does", "... personal Identity can by us be placed in nothing but consciousness (which is that alone which makes what we call self)..."
5. Cf. the central passage: "For since consciousness always accompanies thinking, and it is that, that makes every one to be, what he calls *self*; and thereby distinguishes himself from all other thinking things, in this alone consists *personal Identity*, i.e. the sameness of a rational Being: And as far as this consciousness can be extended backwards to any past Action or Thought, so far reaches the Identity of that *Person*" (Locke 1700/1799:9; cf. also 14: "... personal Identity reaching no farther than consciousness reaches"). The criterion of personal identity then is whether the individual in question can "repeat the *Idea* of any past Action with the same consciousness it had of it at first, and with the same consciousness it has of any present Action; so far it is the same *personal self*" (Locke 1700/1799:10).
6. The article reduces "the problem of the unity of a person through change" (ibid.) to the problem of criteria of identity. As there is no need for a hidden substance for the retention of a person's identity, "all that seems to remain is the much harder question of what changes are allowed for in this concept – the problem of the criteria of identity" (Penulham 1967:96). Thus, the issue is to know what alterations are and are not allowed for: "A man can change in more ways before he is destroyed than a chair can" (Penulham 1967:99). In the following, I will argue that the problem of the unity through time is not a matter of identity of a substance, but part of what it is to be a person.
7. Cf. Heidegger (1927/1972: §31): "Dasein als Verstehen."
8. For Locke, self or person is neither immaterial substance (soul), nor material substance (body). Locke's "wonderful move" (to use Simon Blackburn's expression (1999:128)) is the following: If we would place ourselves in some immaterial substance, we would have no idea of what this substance would be (if we had, this substance would be ourselves: *that* we remember, etc.). Still, for Locke, there is a substance, both immaterial and material, for being

a person. A human being (man) is soul and body. Thus, personhood is related to substances, but in a contingent way (man is not as such a person). For Kierkegaard, in contrast, a human being is defined as a *synthesis* of body and soul united in spirit. Spirit means self: A human being thus relates to itself *as* body and *as* soul (cf. Grøn 1999).

9. In his dissertation *Of Personal Identity*, quoted in Locke (1959: 449f., Note 4).

10. Cf., e.g.: “Had I the same consciousness, that I saw the Ark and *Noah’s* flood, as that I saw an overflowing of the *Thames* last Winter, or as that I write now, I could no more doubt that I, that write this now, that saw the *Thames* overflowed last Winter, and that viewed the Flood at the general Deluge, was the same *self*, place that *self* in what Substance you please, than that I that write this am the same *my self* now whilst I write (whether I consist of all the same Substance, material or immaterial, or no) that I was Yesterday” (Locke 1700/1979: 16).

11. Cf.: “For it is by the consciousness it [an intelligent being] has of its present Thoughts and Actions, that it is *self* to it *self* now, and so will be the same *self* as far as the same consciousness can extend to Actions past or to come” (Locke 1700/1979: 10). This poses a problem, that is, to account for the claimed fact that I can *now* remember how I was aware of myself yesterday or 10 years ago. That I *can* now remember myself doing what I did yesterday or 10 years ago could be taken as support for the second interpretation (memory as self-awareness).

12. The issue could also be the following: Am I the same by having the same memory, or is memory consciousness of being the same?

13. This could be taken as a further argument for the second interpretation of memory as self-awareness: What I remember is not only what I have done and thought, but also myself having done and thought so.

14. In Locke, references to God (Locke 1700/1979: 13) and to “the great Day” are not disquieting, but quite reassuring. In “the great Day”, there is still no problem about consciousness: “But in the great Day, wherein the Secrets of all Hearts shall be laid open, it may be reasonable to think, no one shall be made to answer for what he knows nothing of; but shall receive his Doom, his Conscience accusing or excusing him” (Locke 1700/1979: 22). Maybe already in the present day of remembering, conscience might disturb our consciousness.

15. For Locke, to be concerned has to do with one’s own happiness and misery. In the case before us, this criterion is both relevant and would lose the narrow interpretation given to it by Locke.

16. The relation between self and body then is no longer enigmatic but contingent. This opens up puzzling cases such as, for example, “the Soul of a Prince, carrying with it the consciousness of the Prince’s past Life, enter and inform the Body of a Cöbler” (Locke 1700/1979: 15). If I imagine that I could have my soul in different places and times, and different bodies from my own, what then would my soul be? It would be some thing, not me, asking this question.

17. Cf. Locke (1700/1979: 14): Can I conceive myself the same person with Nestor or Thersites? Can I find myself the same person with Nestor?

18. Perronet, *Vindication of Locke*, quoted in Locke (1959: 454, Note 3).

19. "Nul ne peut traiter un homme 'comme un chien', s'il ne le tient d'abord pour un homme. L'impossible déshumanisation de l'opprimé se retourne et devient l'aliénation de l'opprimeur: c'est lui-même qui ressuscite par son moindre geste l'humanité qu'il veut détruire" (Sartre 1957:313).
20. I can only ascribe personal predicates to myself if I also see others as self-ascribers (cf. Strawson 1959: 108). However, I should see others, not only as self-ascribers, but also as *others*-ascribers, as myself. The problem is that I do not simply see others as other selves, both in the sense of self-ascribers and others-ascribers. It might be that I see others without realizing what it means that they are other selves. I might see them in ways in which I ignore that they also ascribe the same predicates to themselves as I do to myself, and that they see me as I see them. What I overlook, then, might also be what I presuppose in taking others to be able to respond to what I do (cf. the quotation above from Sartre). This negative possibility (of ignoring or overlooking) points to the normative dimension of self-identification in relation to others. In this sense, the speaker-hearer situation is not only a matter of identifying whether we are talking about the same thing, but also of understanding each other.
21. Locke (1700/1979:20) refers to both phrases, but in a third person perspective, in the case of madness.
22. Kierkegaard (1849/1980). Cf. Theunissen (1981, 1993) and Grøn (1996a, 1996b).
23. The quotes are taken from a paper on structural disorders in schizophrenia by Josef Parnas.
24. One should then look more carefully into the reflective situation, as a form of speaker-hearer situation. It is reflective in the sense that the context that was implicitly understood can come to the fore.
25. Cf. e.g. Nietzsche (1988:398 ("Das "Subjekt" ist ja nur eine Fiktion"), 162, 315 and 383).
26. We are always "putting the best 'faces'" on ourselves, as Daniel C. Dennett says (Dennett 1992).
27. This study has been funded by the Danish National Research Foundation.

References

- Blackburn, S. (1999). *Think*. Oxford: Oxford University Press.
- Dennett, D. C. (1992). The self as a center of narrative gravity. In F. Kesselet et al. (Eds.), *Self and consciousness: Multiple perspectives*. Hillsdale, NJ: Erlbaum.
- Gallagher, S. (2000). Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences*, 4, 14–21.
- Garrett, B. (2001). Personal identity. *Routledge encyclopedia of philosophy*. CD-Rom. General Editor E. Craig.
- Grøn, A. (1996a). Der Begriff Verzweiflung. In *Kierkegaard Studies. Yearbook 1996* (pp. 33–60). Berlin/New York: de Gruyter.
- Grøn, A. (1996b). Kierkegaards Phänomenologie? In *Kierkegaard Studies. Yearbook 1996* (pp. 91–116). Berlin/New York: de Gruyter.

- Grøn, A. (1999). *Angst bei Søren Kierkegaard. Eine Einführung in sein Denken*. München: Klett-Cotta.
- Hegel, G. W. F. (1807/1952). *Phänomenologie des Geistes*. Hamburg: Felix Meiner.
- Heidegger, M. (1927/1972). *Sein und Zeit*. Tübingen: Niemeyer.
- Henrich, D. (1979). "Identität" – Begriffe, Probleme, Grenzen. In O. Marquard & K. Stierle (Eds.), *Identität. Poetik und Hermeneutik VIII* (pp. 133–186). München: Wilhelm Fink.
- Hume, D. (1739/1896). *A treatise on human nature*. L. A. Selby-Bigge (Ed.). Oxford: Oxford University Press.
- Kierkegaard, S. (1849/1980). *The sickness unto death. Kierkegaard's writings*, Vol. 19. Howard V. Hong & Edna H. Hong (Eds., Trans.). Princeton, NJ: Princeton University Press.
- Locke, J. (1700/1979). *An essay concerning human understanding*. Edited with a foreword by P. H. Nidditch. Oxford: Oxford University Press (based on the fourth edition of the *Essay*).
- Locke, J. (1959). *An essay concerning human understanding*, Vol. 1. Collated and annotated by A. C. Fraser. New York: Dover Publications.
- Nietzsche, F. (1988). *Kritische Studienausgabe. Vol. 12. Nachlaß 1885–1887*. Herausgegeben von Giorgio Colli und Mazzino Montinari. Berlin/New York: de Gruyter.
- Penulham, T. (1967). Personal identity. In P. Edwards (Ed.), *The encyclopedia of philosophy*, Vol. 6 (pp. 95–107). New York/London: Macmillan.
- Ricoeur, P. (1992). *Oneself as another*. Chicago & London: The University of Chicago Press.
- Sartre, J.-P. (1957/1970). *Les temps modernes* (1957), quoted from Contat, M. & Rybalka, M.: *Les Écrits de Sartre*, Paris 1970.
- Strawson, P. F. (1959/1964). *Individuals. An essay in descriptive metaphysics*. London: Methuen.
- Taylor, C. (1989). *Sources of the self. The making of modern identity*. Cambridge: Cambridge University Press.
- Theunissen, M. (1981). Kierkegaard's negativistic method. *Psychiatry and the humanities*, Vol. 5 (pp. 381–423). New Haven: Yale University Press.
- Theunissen, M. (1993). *Der Begriff Verzweiflung. Korrekturen an Kierkegaard*. Frankfurt am Main: Suhrkamp.
- Williams, B. (1973/1999). *Problems of the self*. Cambridge: Cambridge University Press.
- Zahavi, D. (2003). Phenomenology of self. In T. Kircher & A. David (Eds.), *The self in neuroscience and psychiatry* (pp. 56–75). Cambridge: Cambridge University Press.

Index

A

Adamson, L. B. 22, 29
Agnetta, B. 15
Anisfeld, M. 8
Austin, J. L. 30
Avramides, A. 50, 59

B

Baddeley, A. D. 24
Bahrack, L. 9, 13, 67
Bakeman, R. 22
Baker, L. R. IX, X
Baron-Cohen, S. 36, 37, 41, 59
Bates, E. 10
Beckett, S. 120
Bergman, A. 6, 43
Bergman, T. 7
Berman, K. F. 83
Bermúdez, J. L. 58
Bertenthal, B. 2, 6, 9
Binet, A. 79
Blackburn, S. 37
Blakemore, S. 77, 83, 92, 101, 102
Boccia, M. L. 10
Boks, P. 105
Botvinick, M. 66
Bowlby, J. 22
Brazelton, T. B. 29
Brentano, F. 60
Bretherton, I. 22, 24, 25, 32
Brissenden, S. 106
Brooks-Gunn, J. 2, 4, 16, 46
Bruner, J. S. 30
Bull, A. 118, 119
Busnell, M. C. 6

Butler, J. 132

Butterworth, G. E. 6, 21, 30, 43–45,
52, 53, 55, 56, 58

C

Callaghan, T. C. 12
Cambell, R. 12
Campbell, J. 94, 95
Campos, J. J. 24
Carpenter, E. 2
Carruthers, P. 36, 38, 40–42, 59
DeCasper, A. J. 6
Chaminade, T. 102
Chua, H. C. 100
Cloninger, C. R. 101
Cohen, J. 66
Cole, J. XIII, 105, 106
Copernicus, N. 59
Corboz-Warnery, A. 23, 29
Cowan, P. 23

D

Damasio, A. 105
Daprati, E. 73–75, 83, 97
Darwiche, J. 25
Davidson, D. 49, 50
Decety, J. 79, 80, 102
Della Sala, S. 100
DeLoache, J. S. 12
Dennett, D. C. 93, 155
De Noni, S. 29
De Vries, J. I. P. 6
Dierks, T. 84
Duckworth, S. 110–112, 114, 117
Dunn, J. 22

E

Emde, R. N. 24

F

Faidil, C. 9

Fantz, R. L. 9

Farné, A. 66

Farrer, C. 74, 97, 102

Favez, N. XI, 29

Feinberg, I. 78

Fifer, W. P. 6

Fink, G. R. 102

Finkelstein, V. 113

Fisher, K. 2, 9

Fivaz-Depeursinge, E. XI, 23, 25, 29,
32, 47

Flach, R. 80

Flanagan, O. 60, 93

Fontaine, A.-M. 16

Fourneret, P. 73, 102

Frak, V. G. 79

Franck, N. 83, 84, 97, 101

Frank, M. 60

Frankfurt, H. 90

Frascarolo, F. XI, 25, 29

Frith, C. D. 81, 92, 94, 95, 97,
100–102

Frith, U. 38, 41, 59

Frye, D. 15

Fulford, K. W. M. 92

G

Gallagher, S. XII, XIII, 26, 31, 32, 58,
66, 91, 97, 98, 100–102, 105, 148

Gallese, V. 80

Gallup, G. G. 16

Garrett, B. 126

Georgieff, N. 82, 96, 102

Gesinus-Visser, G. 32

Gibson, J. J. 44

Goldman, A. 80

Goldman, S. 119

Goldman-Rakic, P. S. 83

Goldstein, K. 109

Gopnik, A. 36, 38–41, 59

Gordon, R. M. 36

Goubet, N. 15

Graham, G. XII, XIII, 91, 93, 94, 98,
99, 101

Grapenne, O. 6

Grøn, A. XIII, 154, 155

H

Haggard, P. 77

Haith, M. M. 7

Happé, F. 38, 41, 59

Harris, P. 13, 66

Heal, J. 36

Hegel, G. W. F. 139, 140

Heidegger, M. 153

Henrich, D. 60, 153

Henry, M. 57, 60

Hespos, S. J. 5

Hesslow, G. 78

Hobson, R. P. 49, 50

Hoffman, R. 94

Hopkins, B. 6

Hubley, P. 21, 22

Hume, D. 151

Husserl, E. 57–60

I

Inhelder, B. 43

J

Jackson, M. C. 92

James, W. 16

Jeannerod, M. XII, 72–75, 78–80,
82, 96, 97, 102

Johnson, M. H. 7

Johnson, S. H. 79

Jouen, F. 6

K

Kahn, R. L. 72

Kepler, J. 59

Kierkegaard, S. 130, 131, 138, 144,
149, 154, 155

Klennert, M. D. 24
 Klosterkötter, J. 96
 Knoblich, G. 75, 80
 Kriegel, U. 60

L

Lavanchy, C. 29
 Leblond, C. 15
 Lecanuet, J. P. 6
 Leder, D. 60, 105
 Legerstee, M. 47
 Leslie, A. M. 37
 Lewis, M. 2, 4, 13, 16, 46
 Libet, B. 77
 Lipps, T. 80
 Locke, J. 125, 131–137, 142, 143,
 145, 146, 153–155

M

Mahler, M. S. XI, 6, 43
 Marcel, A. 31
 Marchetti, C. 100
 Marlier, L. 6
 Maurer, D. 7
 McCulloch, G. 49
 McHale, J. 23
 Mead, G. H. X
 Meltzoff, A. N. 7, 8, 15, 21, 26, 27,
 31, 36
 Merleau-Ponty, M. 49–52, 57, 59,
 110, 115, 120, 121
 Mitchell, R. W. 10
 Mittelstaedt, H. 76
 Moore, M. J. 7
 Moore, M. K. 7, 8, 21, 26, 27, 31
 Morris, J. 117
 Morton, J. 7
 Moss, L. 9
 Mullins, S. 91, 95, 96
 Murphy, R. F. 106, 110, 115, 119

N

Nadel, J. 31
 Neisser, U. 43–45, 52, 53, 56, 58

Nichols, S. 41
 Nielsen, T. I. 68, 70–73, 84
 Nietzsche, F. 151, 152, 155
 Nöe, A. 101

O

Oakeley, D. A. 101
 Ohlson, D. 12
 Oliver, M. 110, 112–114

P

Paccalin, C. 82
 Papousek, H. 27
 Papousek, M. 27
 Parker, S. T. 10
 Parnas, J. 60, 96, 155
 Parsons, L. M. 79
 Penulham, T. 126, 153
 Perner, J. 12
 Piaget, J. XI, 6, 10, 11, 14, 15, 43
 Pine, F. 6, 43
 Poss, S. 15
 Povinelli, D. J. 3, 9, 11, 12
 Precht, H. F. R. 6
 Premack, D. 35
 Prinz, W. 80

R

Ramachandran, V. S. 66
 Reddy, V. 16
 Rimbaud, A. 1
 Ricoeur, P. 142, 147
 Rochat, P. XI, XIII, 3, 5, 7–9, 11–13,
 15, 21, 22, 30–32, 43–46, 48, 52,
 53, 55, 56, 58
 Rogers-Ramachandran, D. 66
 Rose, J. L. 6
 Ruby, P. 102

S

Sacks, O. 105
 Salapatek, P. 7
 Sartre, J.-P. 57, 59, 60, 140, 155

Sass, L. 100, 102
Schaal, B. 6
Schaffer, H. R. 22
Schneider, K. 83
Selemon, L. D. 83
Shoemaker, S. 54, 55, 106
Sirigu, A. 79
Slachevsky, A. 7
Smith, P. K. 38
Sorce, J. F. 24
Soussignan, R. 6
Spence, S. A. 84, 91, 95, 96, 97
Sperry, R. W. 76
Spitz, R. A. 47
Stephens, G. L. XII, XIII, 91, 93, 94,
98, 99, 101
Stern, D. 22, 23, 25, 30, 43, 44,
46–48, 52–56, 58
Stich, S. 41
Stover, S. L. 117
Strawson, G. 60
Strawson, P. F. 50, 153, 155
Striano, T. 8, 9, 13, 21, 22, 30
Sullivan, R. 71, 72, 84
Svejda, M. 24
Sørensen, J. B. 84

T

Taylor, C. 135
Theunissen, M. 155
Thompson, E. 101
Tomasello, M. 13, 21
Tow, A. M. 100
Tremblay-Leveau, H. 31

Trevarthen, C. 21, 22, 47
Tronick, E. 29

V

van den Bos, E. 75, 102
Varela, F. J. 105
Vasey, S. 118
Visser, G. H. A. 6
Vogele, K. 97, 102
Von Holst, E. 76

W

Waldenfels, B. 60
Watson, J. S. 67
Weinberger, D. R. 83
Weinstein, E. A. 72
Wegner, D. 66, 78
Wellman, H. M. 59
Weiskrantz, L. 105
Wilkes, K. V. 60
Williams, B. 135
Wise, S. 29
Wittgenstein, L. 51, 58, 59
Wolff, P. 7
Wolpert, D. M. 77, 92, 101
Woodruff, G. 35

Z

Zaner, R. M. 60
Zahavi, D. XII, XIII, 32, 60, 96, 147
Zazzo, R. 3, 4, 16
Zola, I. 112

In the series *Advances in Consciousness Research* the following titles have been published thus far or are scheduled for publication:

- 1 **GLOBUS, Gordon G.:** *The Postmodern Brain*. 1995. xii, 188 pp.
- 2 **ELLIS, Ralph D.:** *Questioning Consciousness. The interplay of imagery, cognition, and emotion in the human brain*. 1995. viii, 262 pp.
- 3 **JIBU, Mari and Kunio YASUE:** *Quantum Brain Dynamics and Consciousness. An introduction*. 1995. xvi, 244 pp.
- 4 **HARDCASTLE, Valerie Gray:** *Locating Consciousness*. 1995. xviii, 266 pp.
- 5 **STUBENBERG, Leopold:** *Consciousness and Qualia*. 1998. x, 368 pp.
- 6 **GENNARO, Rocco J.:** *Consciousness and Self-Consciousness. A defense of the higher-order thought theory of consciousness*. 1996. x, 220 pp.
- 7 **MAC CORMAC, Earl and Maxim I. STAMENOV (eds.):** *Fractals of Brain, Fractals of Mind. In search of a symmetry bond*. 1996. x, 359 pp.
- 8 **GROSSENBACHER, Peter G. (ed.):** *Finding Consciousness in the Brain. A neurocognitive approach*. 2001. xvi, 326 pp.
- 9 **Ó NUALLÁIN, Seán, Paul Mc KEVITT and Eoghan Mac AOGÁIN (eds.):** *Two Sciences of Mind. Readings in cognitive science and consciousness*. 1997. xii, 490 pp.
- 10 **NEWTON, Natika:** *Foundations of Understanding*. 1996. x, 211 pp.
- 11 **PYLKKÖ, Pauli:** *The Aconceptual Mind. Heideggerian themes in holistic naturalism*. 1998. xxvi, 297 pp.
- 12 **STAMENOV, Maxim I. (ed.):** *Language Structure, Discourse and the Access to Consciousness*. 1997. xii, 364 pp.
- 13 **VELMANS, Max (ed.):** *Investigating Phenomenal Consciousness. New methodologies and maps*. 2000. xii, 381 pp.
- 14 **SHEETS-JOHNSTONE, Maxine:** *The Primacy of Movement*. 1999. xxxiv, 584 pp.
- 15 **CHALLIS, Bradford H. and Boris M. VELICHKOVSKY (eds.):** *Stratification in Cognition and Consciousness*. 1999. viii, 293 pp.
- 16 **ELLIS, Ralph D. and Natika NEWTON (eds.):** *The Caldron of Consciousness. Motivation, affect and self-organization — An anthology*. 2000. xxii, 276 pp.
- 17 **HUTTO, Daniel D.:** *The Presence of Mind*. 1999. xiv, 252 pp.
- 18 **PALMER, Gary B. and Debra J. OCCHI (eds.):** *Languages of Sentiment. Cultural constructions of emotional substrates*. 1999. vi, 272 pp.
- 19 **DAUTENHAHN, Kerstin (ed.):** *Human Cognition and Social Agent Technology*. 2000. xxiv, 448 pp.
- 20 **KUNZENDORF, Robert G. and Benjamin WALLACE (eds.):** *Individual Differences in Conscious Experience*. 2000. xii, 412 pp.
- 21 **HUTTO, Daniel D.:** *Beyond Physicalism*. 2000. xvi, 306 pp.
- 22 **ROSSETTI, Yves and Antti REVONSUO (eds.):** *Beyond Dissociation. Interaction between dissociated implicit and explicit processing*. 2000. x, 372 pp.
- 23 **ZAHAVI, Dan (ed.):** *Exploring the Self. Philosophical and psychopathological perspectives on self-experience*. 2000. viii, 301 pp.
- 24 **ROVEE-COLLIER, Carolyn, Harlene HAYNE and Michael COLOMBO:** *The Development of Implicit and Explicit Memory*. 2000. x, 324 pp.
- 25 **BACHMANN, Talis:** *Microgenetic Approach to the Conscious Mind*. 2000. xiv, 300 pp.
- 26 **Ó NUALLÁIN, Seán (ed.):** *Spatial Cognition. Foundations and applications*. 2000. xvi, 366 pp.
- 27 **GILLET, Grant R. and John McMILLAN:** *Consciousness and Intentionality*. 2001. x, 265 pp.
- 28 **ZACHAR, Peter:** *Psychological Concepts and Biological Psychiatry. A philosophical analysis*. 2000. xx, 342 pp.
- 29 **VAN LOOCKE, Philip (ed.):** *The Physical Nature of Consciousness*. 2001. viii, 321 pp.
- 30 **BROOK, Andrew and Richard C. DEVIDI (eds.):** *Self-Reference and Self-Awareness*. 2001. viii, 277 pp.

- 31 **RAKOVER, Sam S. and Baruch CAHLON:** Face Recognition. Cognitive and computational processes. 2001. x, 306 pp.
- 32 **VITIELLO, Giuseppe:** My Double Unveiled. The dissipative quantum model of brain. 2001. xvi, 163 pp.
- 33 **YASUE, Kunio, Mari JIBU and Tarcisio DELLA SENTA (eds.):** No Matter, Never Mind. Proceedings of Toward a Science of Consciousness: Fundamental approaches, Tokyo 1999. 2002. xvi, 391 pp.
- 34 **FETZER, James H. (ed.):** Consciousness Evolving. 2002. xx, 253 pp.
- 35 **Mc KEVITT, Paul, Seán Ó NUALLÁIN and Conn MULVIHILL (eds.):** Language, Vision and Music. Selected papers from the 8th International Workshop on the Cognitive Science of Natural Language Processing, Galway, 1999. 2002. xii, 433 pp.
- 36 **PERRY, Elaine, Heather ASHTON and Allan H. YOUNG (eds.):** Neurochemistry of Consciousness. Neurotransmitters in mind. With a foreword by Susan Greenfield. 2002. xii, 344 pp.
- 37 **PYLKKÄNEN, Paavo and Tere VADÉN (eds.):** Dimensions of Conscious Experience. 2001. xiv, 209 pp.
- 38 **SALZARULO, Piero and Gianluca FICCA (eds.):** Awakening and Sleep–Wake Cycle Across Development. 2002. vi, 283 pp.
- 39 **BARTSCH, Renate:** Consciousness Emerging. The dynamics of perception, imagination, action, memory, thought, and language. 2002. x, 258 pp.
- 40 **MANDLER, George:** Consciousness Recovered. Psychological functions and origins of conscious thought. 2002. xii, 142 pp.
- 41 **ALBERTAZZI, Liliana (ed.):** Unfolding Perceptual Continua. 2002. vi, 296 pp.
- 42 **STAMENOV, Maxim I. and Vittorio GALLESE (eds.):** Mirror Neurons and the Evolution of Brain and Language. 2002. viii, 392 pp.
- 43 **DEPRAZ, Nathalie, Francisco J. VARELA and Pierre VERMERSCH:** On Becoming Aware. A pragmatics of experiencing. 2003. viii, 283 pp.
- 44 **MOORE, Simon C. and Mike OAKSFORD (eds.):** Emotional Cognition. From brain to behaviour. 2002. vi, 350 pp.
- 45 **DOKIC, Jérôme and Joëlle PROUST (eds.):** Simulation and Knowledge of Action. 2002. xxii, 271 pp.
- 46 **MATEAS, Michael and Phoebe SENGERS (eds.):** Narrative Intelligence. 2003. viii, 342 pp.
- 47 **COOK, Norman D.:** Tone of Voice and Mind. The connections between intonation, emotion, cognition and consciousness. 2002. x, 293 pp.
- 48 **JIMÉNEZ, Luis (ed.):** Attention and Implicit Learning. 2003. x, 385 pp.
- 49 **OSAKA, Naoyuki (ed.):** Neural Basis of Consciousness. 2003. viii, 227 pp.
- 50 **GLOBUS, Gordon G.:** Quantum Closures and Disclosures. Thinking-together postphenomenology and quantum brain dynamics. 2003. xxii, 200 pp.
- 51 **DROEGE, Paula:** Caging the Beast. A theory of sensory consciousness. 2003. x, 183 pp.
- 52 **NORTHOFF, Georg:** Philosophy of the Brain. The brain problem. 2004. x, 433 pp.
- 53 **HATWELL, Yvette, Arlette STRERI and Edouard GENTAZ (eds.):** Touching for Knowing. Cognitive psychology of haptic manual perception. 2003. x, 322 pp.
- 54 **BEAUREGARD, Mario (ed.):** Consciousness, Emotional Self-Regulation and the Brain. 2004. xii, 294 pp.
- 55 **PERUZZI, Alberto (ed.):** Mind and Causality. 2004. xiv, 235 pp.
- 56 **GENNARO, Rocco J. (ed.):** Higher-Order Theories of Consciousness. An Anthology. 2004. xii, 371 pp.
- 57 **WILDGEN, Wolfgang:** The Evolution of Human Language. Scenarios, principles, and cultural dynamics. 2004. x, 237 pp.
- 58 **GLOBUS, Gordon G., Karl H. PRIBRAM and Giuseppe VITIELLO (eds.):** Brain and Being. At the boundary between science, philosophy, language and arts. xii, 350 pp.