



Åbo Akademi
University

Ylva Gustafsson

Interpersonal Understanding and Theory of Mind

Åbo Akademi University Press
Tavastgatan 13, FI-20500 Åbo, Finland
Tel. +358 (0)2 215 3478
E-mail: forlaget@abo.fi

Sales and distribution:
Åbo Akademi University Library
Domkyrkogatan 2–4, FI-20500 Åbo, Finland
Tel. +358 (0)2 -215 4190
E-mail: publikationer@abo.fi

INTERPERSONAL UNDERSTANDING AND THEORY OF MIND



Interpersonal Understanding and Theory of Mind

Ylva Gustafsson

Åbo Akademis förlag | Åbo Akademi University Press
Åbo, Finland, 2014

CIP Cataloguing in Publication

Gustafsson, Ylva.

Interpersonal understanding and
theory of mind / Ylva Gustafsson. -
Åbo : Åbo Akademi University Press,
2014.

Diss.: Åbo Akademi University.
ISBN 978-951-765-748-8

ISBN 978-951-765-748-8
ISBN 978-951-765-749-5 (digital)
Painosalama Oy
Åbo 2014

Table of Contents

ACKNOWLEDGEMENTS	I
INTRODUCTION	1
<i>Two theoretical orientations within theory of mind</i>	5
1. <i>Emotivistic theory-of-mind theories</i>	7
2. <i>Rationalistic theory-of-mind theories</i>	14
<i>Some general theoretical problems with theory-of-mind theories</i>	17
<i>The relationship between theory-of-mind theories and empirical research</i>	25
<i>Summaries of the chapters</i>	36
CHAPTER 1: RATIONALISTIC AND EMOTIVISTIC THEORIES OF ALTRUISM.....	41
1.1 <i>Introduction</i>	41
1.2 <i>Altruism as transaction</i>	43
1.3 <i>Emotivistic conceptions of altruism</i>	47
1.4 <i>Sympathy, empathy and retributive emotions as consisting of analogical imagination</i>	53
1.5 <i>Criticism of the argument from analogy</i>	60
1.6 <i>Cooperation: emotivistic conceptions of human social life as instrumental</i>	66
1.7 <i>Transaction: a rationalistic conception of human social life as instrumental</i>	76
1.8 <i>Conclusion</i>	84
CHAPTER 2: THEORY OF MIND AND INFANTS' IMITATION OF FACIAL EXPRESSIONS	88
2.1 <i>Introduction</i>	88
2.2 <i>Empirical research on infants' imitation of facial expressions</i>	89

2.3 <i>The influence from natural science on empirical research methods</i>	97
2.4 <i>The influence from theory-of-mind theory on empirical research methods</i>	100
2.5 <i>What do we mean by imitation</i>	110
2.6 <i>A dialogical perspective</i>	114
2.7 <i>Mirror neurons</i>	123
2.8 <i>Sharing experiences</i>	128
2.9 <i>Conclusion</i>	138
CHAPTER 3: AUTISM AND THEORY OF MIND	141
3.1 <i>Introduction</i>	141
3.2 <i>Autism as “mindblindness”</i>	143
3.3 <i>Autism, pretend play and mindreading</i>	154
3.4 <i>Autism and the false belief task</i>	170
3.5 <i>Theory of mind and autistic children’s language problems</i>	184
3.6 <i>Theory of mind and autistic persons’ difficulties to understand irony</i>	194
3.7 <i>Conclusion</i>	202
CHAPTER 4: THE RELATIONSHIP BETWEEN EMPATHIC IMAGINATION AND COMPASSION	203
4.1 <i>Introduction</i>	203
4.2 <i>The relation between empathy and compassion</i>	206
4.3 <i>Stotland’s experimental research on empathy</i>	210
4.4 <i>Batson’s experimental research on empathy</i>	220
4.5 <i>Cognitive conceptions of empathy and conceptual confusion</i>	229
4.6 <i>Imagination, involvement and responsibility</i>	236
4.7 <i>Solipsism, body-mind dualism and experiences of suffering</i>	247

4.8 <i>Imagination as a lack of understanding</i>	264
4.9 <i>Conclusion</i>	275
CONCLUDING REMARKS.....	278
SWEDISH SUMMARY - SAMMANFATTNING	291
BIBLIOGRAPHY.....	296

Acknowledgements

During the work on this thesis I have studied at the Department of Philosophy at Åbo Akademi University. Throughout these years, I have had the fortune to get to know many very good philosophers, which have given me great help with the work on this thesis. During my years at the department, I have also met many good friends.

First and foremost I wish to thank my supervisor, Professor emeritus Lars Hertzberg, who has patiently helped me along philosophically. His insightful and sensitive philosophical reflections have been of an enormous help. Lars' philosophical thinking and writing has influenced me extensively, and I am grateful for having had the privilege to have him as teacher.

Many people have given me philosophical inspiration and invaluable help. I want to thank Professor Martin Gustafsson, docent Olli Lagerspetz and artist and philosopher Göran Torrkulla. Martin, who succeeded Lars some years ago, has offered me great help. His encouraging and constructive comments have been very important. Olli and Göran have both been my teachers since I started to study philosophy twenty years ago. Their philosophical reflections during courses and in the research seminars have been of great importance for me.

I am also grateful to Lars Hertzberg, Olli Lagerspetz and docent Petter Korkman for giving me the opportunity to be part of three philosophical projects. These projects have provided several years of financial support as well as interesting cooperation with other researchers.

Camilla Kronqvist and I have worked together in all these three projects. Our close philosophical perspective as well as our cooperation for many years has been very rewarding.

Mari Lindman and Camilla have also been my roommates for many years. I have very much enjoyed their daily company. We have had many fun and intriguing discussions about philosophy and life.

I further want to thank my colleagues Jonas Ahlskog, Benjamin Alm, Joel Backström, Ann-Helen Berg, Kim Berts, Niklas Forsberg, Antony Fredriksson, Yrsa Neuman, Martin Nybom, Hannes Nykänen, Åsa Slotte, Hugo Strandberg and Iris Wikström. I have greatly appreciated their philosophical reflections and enjoyed our lunches, coffee breaks and beers with discussions about all kinds of topics. My warm thanks also to former colleagues, and to all the participants at the research seminar.

The historian Anders Ahlbäck has been a dear and supportive friend of our family during all these years. He has also given me helpful interdisciplinary comments on the thesis.

My parents, Picke and Classe, have always encouraged me in my studies. I have enjoyed many talks with them about life and about philosophy. My mother has been of an enormous practical help. She has helped me endless amounts of times with taking care of my children.

My partner Benjamin has been part of my life since we both started studying philosophy twenty years ago. Many of the things we have done together have been great fun. However, it is the life with our children that has been the

greatest adventure we have shared. Our children Ville and Maja were born when I had just started the work on the thesis. They give me great joy.

Finally, I want to express my gratitude for the large financial support I have received in order to work on the thesis, by the following projects and funds: The Ella and Georg Ehrnrooth foundation, The Finnish Academy of Science project *Emotions in Dialogue, Perspectives from the Humanities* led by docent Olli Lagerspetz, The Finnish Academy of Science project *Emotions and Understanding* led by Professor Lars Hertzberg, The Finnish Konkordia fund, The Waldemar von Frenckell foundation, The Kone foundation project *Westermarck and Beyond, Evolutionary Approaches to Morality and Their Critics* led by docent Olli Lagerspetz, The Edward Westermarck foundation, The Vice-Chancellor of Åbo Akademi, The Åbo Akademi University Endowment Research Institute, The Oskar Öflund foundation.

Introduction

What does it mean to understand someone else? At the centre of contemporary discourse regarding interpersonal understanding, we find the conception of an individual's *theory of mind*: a theory of the mind of the other. The aim of this thesis is to scrutinise this discourse—what we might call *the philosophy of theories of mind*.¹ The claim that a theory of mind, or, in other words, “mindreading” or “mentalising”, is a fundamental capacity that grounds human social life is popular within both modern philosophical and psychological theorising on interpersonal understanding. This claim also surfaces in evolutionary psychology, in theories of child development, in theories of autism as well as in philosophy on emotions and in moral philosophy. For instance, the psychologists Helen L. Gallagher and Christopher D. Frith (2003) write:

One aspect of social cognition sets us apart from other primates. It underpins our ability to deceive, cooperate and empathize, and to read others' body language. It also enables us to accurately anticipate other people's behaviour, almost as if we had read their minds. This exceptional capacity is known as having a 'theory of mind', or mentalizing. It underlies our ability to explain and predict the behaviour of ourselves and

¹ “Theory of mind” should not be confused with “philosophy of mind”. Philosophy of mind is not the name of a theory but a philosophical field of research, concerned with questions about human or animal understanding. “Theory of mind” is, however, a theoretical perspective on interpersonal understanding.

others by attributing to them independent mental states, such as beliefs, desires, emotions or intentions. (Gallagher and Frith 2003, p. 77)

The anthropologist John Tooby and the psychologist Leda Cosmides (1997) write:

We are 'mindreaders' by nature, building interpretations of the mental events of others and feeling our constructions as sharply as the physical objects we touch. Humans evolved this ability because, as members of an intensively social, cooperative, and competitive species, our ancestors' lives depended on how well they could infer what was on one another's minds. (Tooby and Cosmides 1997, p. xvii)

The philosopher Alvin I. Goldman (2006) writes:

[...] Homo sapiens is a particularly social species, and one of its social characteristics is especially striking: reading one another's minds. People attribute to self and others a host of mental states, ranging from beliefs and aspirations to headaches, disappointments, and fits of anger. [...] Mentalizing may be the root of our elaborate social nature. Would there be language and discourse without mentalizing? Would the exquisitely coordinated enterprises of cultural life, the structures of love, politics, and games, be what they are without participants' attending to the mental states of others? (Goldman 2006, p. 3)

On the surface such descriptions can look all right. As Gallagher and Frith note, is it not a central aspect of human

life that we can understand other people's "body language" and that we can explain and predict the behaviour of other people? As Tooby and Cosmides note, is it not important for human survival that we cooperate with each other? And, as Goldman points out, is it not an important part of our life with each other that we can understand that other people have beliefs, aspirations, headaches, disappointments and fits of anger? However, theory-of-mind theorists do not merely claim that people cooperate or that we can reflect on other people's beliefs, intentions and feelings. They want to make larger claims. According to theory-of-mind theorists, human social life and interpersonal understanding is based on a general mental system or function of "mentalising" or "mindreading". A unifying feature among theory-of-mind theories is the assumption that interpersonal understanding can be regarded as some sort of general cognitive function or method of "mindreading". Another unifying feature among these theories is that it is assumed that we can study these mindreading functions through empirical research. It is often assumed that these functions can be tested by psychological experiments.

Theory-of-mind theories have gained popularity during the last 30 years alongside the increased popularity of cognitive science. Cognitive science has, again, had a large interdisciplinary influence, which is also reflected in the fact that theory-of-mind theories can be found among such various disciplines as philosophy, developmental psychology, neuroscience, evolutionary psychology etc.

While theory-of-mind theories are constantly gaining in popularity there is also an increasing number of critical works directed at theory-of-mind theories.² Among contemporary critical works can, for instance, be mentioned Peter Hobson's *Autism and the Development of Mind* (1993), Dan Zahavi's *Subjectivity and Selfhood* (2005), Vasudevi Reddy's *How Infants Know Minds* (2008), and the anthology *Against Theory of Mind* (2009) edited by Ivan Leudar and Alan Costall. My thesis can be seen as a contribution to this contemporary critical discussion of theory-of-mind theories.

² Theory-of-mind theories are also sometimes called "folk-psychological" theories. When theory-of-mind theories are referred to as folk psychological theories, it is suggested that theory-of-mind theories concern a kind of general mental function or capacity to reason about other people. The theoretical field of folk psychology has been criticized by so called eliminative materialists. According to eliminative materialism our every-day conceptions of beliefs and intentions are based on a flawed, superficial understanding of the minds of other people. Eliminative materialists, such as Paul Churchland, argue that a true understanding of other people ought to be based on neuroscientific findings. I will not discuss the debate on eliminative materialism in this thesis. It is, however, important to note that my criticism of theory of mind in this thesis is not similar to the criticism of folk psychology by eliminative materialists. According to eliminative materialists, proponents of theory of mind (folk psychology) are entangled in conceptual confusions when they talk of people having intentions and beliefs. According to them we ought to avoid such conceptual confusion by instead talking of nerve reactions in the brain. Even though I also argue that theory-of-mind theorists are entangled in conceptual confusion I do not agree with the eliminative materialists. It is in no sense a lesser conceptual confusion if we decide to talk merely of brain states instead of beliefs and intentions.

The aim of this thesis is not to discuss critically the whole theoretical field of theory-of-mind theories. Rather the aim is to scrutinise certain specific influential theory-of-mind theories that are connected with empirical research. Empirical research has a central importance in theory-of-mind theories, giving the theories credibility in various scientific discussions. I argue that the empirical research is often based on certain problematic assumptions about interpersonal relations as well as on various conceptual confusions. However, before saying more about the relation between theory-of-mind theories and their connection to empirical research, I shall describe the varied theoretical field of theory of mind.

Two theoretical orientations within theory of mind

Important to note is that the field of theory-of-mind theory does not consist of *one* theory but of a large number of *different* theories. My use of the concept “theory-of-mind theory” ought therefore to be understood as an umbrella concept comprising a large number of different theories. However, even if there are a large number of different theory-of-mind theories, one can also roughly sort them into two kinds of theoretical orientations. One theoretical approach can broadly be described as emotivistic, while the other can be described as rationalistic. According to the emotivistic theories, interpersonal understanding is dependent on man’s capacity to imagine how another person feels and to feel that emotion himself. The

rationalistic theory-of-mind theories tend, on the other hand, to emphasise reflection and reasoning as the basis for interpersonal understanding. In what follows I shall give a rough outline of these two theoretical fields within theory of mind. Often a distinction is made between *simulation theory* and *theory theory*. This distinction is to a large degree the same as my distinction between emotivistic and rationalistic theories. However, the expressions “simulation theory” and “theory theory” are concerned with a fairly specific contemporary debate in philosophy and psychology. I have therefore chosen to use the broader distinctions “emotivistic” and “rationalistic” theories in this introduction, since I include theorists that would not consider themselves to be part of the specific debate between simulation theory and theory theory. I will, however, refer to simulation theory in chapter two.

It is also important to note that the concept of “theory of mind” is not used in a unitary way. Some researchers use the concept “theory of mind” in a specific sense, meaning a mental capacity (or calculating function) to reason about other minds. Thus, some theorists distinguish between “theory of mind” and “empathy”.³ Some researchers use the concept in a more general sense, meaning a general cognitive function of interpersonal understanding, including emotional reactions. Some of the researchers that I discuss do not use the concept “theory of mind” at all. I use the concept “theory-of-mind theory” in a broad sense,

³ For instance, Andrew Meltzoff (2002) distinguishes between “empathy” and “theory of mind”.

including both emotivistic theories (such as theories on empathy) as well as rationalistic theories on interpersonal understanding, and even though the researchers themselves might not at all use the concept of “theory of mind”.

I should also point out that the theorists that I will discuss in this thesis cannot all be explicitly defined as theory-of-mind theorists. Only some of them would define themselves as such.⁴ Thus it would not be correct to define Martha Nussbaum as a theory-of-mind theorist even if I will maintain that her conception of empathy is reflective of a certain kind of cognitive conception of interpersonal understanding that is common within theory-of-mind theories. Proponents of theory-of-mind are in this sense not a strictly defined group of theorists, neither is the concept of “theory of mind” always used in a unitary way. My aim is then not to specifically discuss only theorists who talk of “theory of mind”. Rather, what I want to do is to discuss certain kind of recurring (problematic) cognitive conceptions of interpersonal understanding and their relation to empirical research.

1. Emotivistic theory-of-mind theories

Within emotivistic theory-of-mind theories a central concept that often comes up is the concept of *empathy*. It is important to note that when philosophers and psychologists talk of empathy or sympathy they often

⁴ Leda Cosmides, John Tooby and also Simon Baron-Cohen might define themselves as theory-of-mind theorists.

mean something different and more specific than we do in ordinary life. In ordinary life when we say that a person is empathic we generally mean that the person cares for other people, that he or she is compassionate and considerate. However, when philosophers and psychologists talk of empathy the meaning is often different. Even if the concept of empathy is often considered to have a connection with compassion it is frequently thought to be a cognitive function or method that enables us to understand other minds and that, as a consequence, can make us inclined to care for others.

There are also many different words used for what broadly can be said to be the same phenomenon. Philosophers and psychologists talk for instance of “simulation”, “sympathy”, “empathy”, “retributive emotions”, “*Einfühlung*” etc. Not only do theorists use different words but they also often mean slightly different things by these words.

The term “empathy” is itself of modern origin. The term was originally coined by the psychologist Edward Titchener (1909). Titchener was himself influenced by the psychologist Theodor Lipps’ (1897) theory about “*Einfühlung*”. Lipps’ theory on *Einfühlung* was originally concerned with aesthetics, discussing optical illusions. Titchener, however, shifted the theoretical focus to psychology. However, even if the concept “empathy” itself is of modern origin, the perspective can be traced back to David Hume’s and Adam Smith’s ideas on sympathy. In *A Treatise of Human Nature* ([1739-40] 1978) Hume writes:

No quality of human nature is more remarkable, both in itself and in its consequences, than that propensity we have to sympathize with others, and to receive by communication their inclinations and sentiments, however different from, or even contrary to our own. [...] When any affection is infus'd by sympathy, it is at first known only by its effects, and by those external signs in the countenance and conversation, which convey an idea of it. This idea is presently converted into an impression, and acquires such a degree of force and vivacity, as to become the very passion itself, and produce an equal emotion, as any original affection. (Hume [1739-40] 1978, pp. 316-317)

According to Hume, human beings have an inclination to “sympathise” with others. This inclination he explains as consisting of a kind of analogical mental function. When I see another person who is happy or sad I will get an idea of the other person’s emotion. This idea will then turn into the emotion itself. Hume also distinguishes between sympathy and compassion. While sympathy is a kind of general mental function that enables us to feel the same as another person, he defines compassion as concern for the other. Still he suggests that compassion is dependent on the mental function of sympathy.

Adam Smith’s conception of sympathy, in *The Theory of Moral Sentiments* ([1759] 2002) largely resembles Hume’s. According to Smith, human beings have a natural inclination to care for other people. This he explains as deriving from a function of our imagination. We care for others because we have a capacity to imagine ourselves in

the other person's situation. Both Hume's and Smith's models can be described as consisting of an argument of analogy. Their interest in sympathy concern moral philosophy, and then, among other things, the origins of compassion. However, their way of explaining compassion is of a cognitive character. Even though Hume and Smith do not use the word empathy but talk of sympathy, their theories of sympathy have much in common with modern theories on empathy.

One theoretical field where Smith's theory of sympathy has been influential is evolutionary psychology. For instance, Charles Darwin's reflections on the "social instinct" in *The Origin of Species and the Descent of Man* ([1859, 1871] 1872) have much in common with Smith's reflections on sympathy. Likewise Edward Westermarck's reflections on "retributive emotions" in *The Origin and Development of the Moral Ideas* (1917) resemble both Smith and Darwin. In modern evolutionary psychology, the primatologist and ethologist Frans de Waal (2009) also reflects on the evolutionary origins of human social life. In quite a similar manner as Darwin and Westermarck, he considers empathy as a basic, natural, emotional-cognitive function that motivates us to care for others. The suggestion is that our care for each other, or in more popular words "altruism", is based on a certain kind of emotional-cognitive function. This function makes us feel and imagine in an analogical sense, thereby making us inclined to care for others.

However, theories of empathy are also often of a more purely cognitive character, and then not concerned with

moral philosophical questions such as compassion. Such cognitively oriented theories are dominant among modern theories on empathy. They are prominent in developmental psychology and in cognitive science. One popular theory in cognitive science is that empathy consists of a brain mechanism. Researchers in neuroscience, such as Giacomo Rizzolatti (2005), have claimed that there are so called “mirror neurons” that enable us to interpret others. Briefly, the theory about mirror neurons consists in the idea that our emotional and bodily responses to other people’s emotions are based on a certain neural system that “mirrors” other people’s emotions and behaviour. It is claimed that this mirror neuron function enables me to feel the same as you feel in a certain situation and thereby enables me to understand you. The “mirror neuron” phenomenon has been studied empirically through brain research on both apes and human beings.

Also in developmental psychology the cognitive focus on empathy has been central. One researcher in this field is the psychologist Andrew Meltzoff. According to Meltzoff (2002), empathy derives from the infant’s capacity to imitate other people’s bodily expressions. Meltzoff has conducted empirical tests with newborn infants, indicating that infants have a natural capacity to imitate other people’s facial expressions. According to Meltzoff, the empirical research indicates that imitation is an important natural mechanism that eventually enhances the child’s capacity for empathy and mindreading.

However, the concept of empathy also surfaces in contemporary philosophical discussions and especially so

in the philosophy on emotion. For instance, Peter Goldie (2000), Martha Nussbaum (2001) and Karsten R. Stueber (2006) discuss empathy. Contemporary philosophical theories on empathy are often also influenced by empirical research. According to Stueber, empathy consists of two levels; *basic empathy* which originates from the mirror neuron function, and *re-enactive empathy* that, according to him, cannot be understood as a brain function but ought to be considered as a matter of reasoning, requiring a capacity to imagine the other person's larger life context, including his life history and his ways of reasoning. Stueber then does not consider empathy merely as a momentary emotional state mirroring another person's current emotional state. In Stueber's theory on empathy one can see how philosophical theory can be influenced by empirical research, such as the research on mirror neurons. At the same time, however, Stueber's conception of empathy has a broader scope than the empirical research, encompassing more reflective capacities to reason than the mirror neuron theory suggests. Peter Goldie's theory on empathy resembles that of Stueber. Also Martha Nussbaum's philosophical reflections on empathy are influenced by empirical research. According to her, there is an empirically observable link between empathic imagination and compassion. In her work she refers to certain psychological experiments concerning the relation between empathy and compassion that have been made by Daniel C. Batson (1991).

In this sense, theoretical discussions concerning empathy span over many theoretical fields (evolutionary

psychology, developmental psychology, philosophy of emotions etc.), and concern many different concepts (sympathy, *Einfühlung*, empathy, retributive emotions, social instincts, mirror neurons, imitation, re-enactive empathy etc.) As one can see from the description above, theories of empathy also concern many different levels. Some talk of brain mechanisms, some talk of a capacity to imitate and decipher bodily expressions, some talk of our emotional responsiveness to others, and some talk of reflective imagination requiring a large acquaintance with the other person's patterns of thinking.

Another difference is that while some researchers who discuss empathy claim that it consists in an effort of analogical imagination, i.e. to imagine *oneself* in the other person's situation, others point out that the other person can be truly *different* from me and that mere analogical imagination will often not enable us to understand the other person. The importance of being aware of the other as having a different perspective is already argued by Smith ([1759] 2002). In contemporary philosophy and psychology it is claimed by, among others, Batson (1991), Goldie (2006), Nussbaum (2001) and Stueber (2006).

Despite many differences in the theories of empathy, there are some central shared assumptions that underlie these theories. As I have already maintained, one central assumption is that interpersonal understanding in some sense works by analogical imagination or by analogical bodily or emotional reactions. Another central idea that unites the theorists on empathy is the idea that human beings are naturally social beings. In this sense empathy

theorists are often critical of an old tendency (both within philosophy as well as within psychology) to assume that man originally is a lonely individual who cares only for himself. Several of these theorists expressly take a stance against such a conception of human life (for instance Smith, de Waal and Meltzoff). A third uniting feature among empathy theorists is that human social life and interpersonal understanding are, in some sense, considered to be based on our spontaneous emotional responsiveness to each other. Here one can again see a critical stance towards certain rationalistic orientations within both philosophy and psychology.

2. Rationalistic theory-of-mind theories

Above I have described a group of theory-of-mind theories that are all in some sense based on a conception of interpersonal understanding as consisting in an emotional-cognitive analogical method or function of imagination. However, there are also theory-of-mind theories that have a more rationalistic character. Often the emotivistic and rationalistic conceptions blend into each other and often researchers embrace both conceptions. Still one might say that some theories have a more rationalistic orientation while other theories are more emotivistic in kind.

A uniting feature of rationalistic theory-of-mind theories is the assumption that we understand other people by using some sort of calculating method or function of reasoning. The rationalistic perspective, however, is often combined with various kinds of

emotivistic theories of interpersonal understanding, including theories of empathy. In this sense the distinction between emotivistic and rationalistic theory-of-mind theories should not be seen as very sharp.

As with the emotivistic theoretical conception, there are also many different theories within the rationalistic conception of theory of mind. One strong branch of rationalistic theory-of-mind theories lies within evolutionary psychology. According to Leda Cosmides and John Tooby (2008), who work within the field of evolutionary psychology, human beings have throughout our evolutionary history developed an *unconscious, algorithmic reasoning mechanism* that enables us to predict and explain other people's behaviour and thoughts. According to them, this mental algorithmic system consists of modules that are designed for various survival purposes. Some of these algorithmic modules are designed to enable the individual to decipher other people's behaviour and thereby predict their actions, enabling the individual to manipulate others. Cosmides and Tooby are influenced by William D. Hamilton's (1964) and John Maynard Smith's (1982) sociobiological theories. The allusion to an algorithmic biological mechanism derives from this influence. However, one can also see an older influence from Thomas Hobbes' ([1651] 1996)⁵ rationalistic political theory of the social contract.

Another strong field in rationalistic theory-of-mind theories concerns research on autism, which can also be

⁵ See, for instance, Hobbes, ([1651] 1996) *Leviathan*, chapters 13–15.

said to be a part of developmental psychology. Children with autism often appear to be unaware of people. This was an observation that was strikingly described already by the psychiatrist Leo Kanner (1943). Kanner describes how the children he studied seemed to be aware of physical objects, playing with toys etc., while they seemed to be almost completely unaware of their parents or other people, not responding to these other people or taking any interest in them. When children with autism grow older they often do gain some understanding of other people, though some remain largely unresponsive to others. Still most people with autism struggle throughout their life with various degrees of difficulties of comprehending other people and of managing in social situations. Autistic persons can also tend to have peculiar problems with language, such as a rigid way of talking, insensitivity to the tone of voice, tendencies for echolalia etc.⁶ According to proponents of theory of mind, the many social and linguistic problems that persons with autism struggle with, reflect a basic inability to see that other people have minds, i.e. to see that people have intentions, thoughts, feelings etc. A major figure in the field of autism research and theory of mind is the psychologist Simon Baron-Cohen. Baron-Cohen (1997) claims that autism derives from certain dysfunctions in the brain, especially a dysfunction in the so called "Theory of Mind Mechanism". According to Baron-Cohen, these dysfunctions lead to "mindblindness" i.e. an

⁶ For a fuller description of autism see: American Psychiatric Association (2013), *Diagnostic and Statistical Manual of Mental Disorders: Fifth Edition*.

inability to theorise about other people's intentions and beliefs. Baron-Cohen's theory on autism is also influenced by Tooby and Cosmides' evolutionary theories on the origins of human social life. It is largely with their theoretical perspective as a background that Baron-Cohen constructs his theory on autism as mindblindness.⁷

I have now tried to give a broad description of the various theories within the field of theory-of-mind theory. I have claimed that one can divide theory-of-mind theories into two theoretical orientations; an emotivistic orientation and a rationalistic orientation. However, I have also maintained that it is important to note that there are many different perspectives *within* these two orientations of theory of mind. It is important to be aware of this diversity, but it is also important to see that this diversity comprises a certain unitary pattern of theoretical thinking.

Some general theoretical problems with theory-of-mind theories

There are certain problematic theoretical assumptions that the theory-of-mind theories share. To begin with, they are all based on the assumption that interpersonal understanding is an epistemological matter. That is, the

⁷ In his book *Mindblindness, an Essay on Autism and Theory of Mind* (1997) Baron-Cohen reflects extensively on the evolutionary origins of interpersonal understanding, referring to research by Tooby and Cosmides. Their close theoretical connection is also reflected in the fact that Tooby and Cosmides have written the foreword to Baron-Cohen's book.

theories are based on the assumption that to understand another person means to have certain information about that person. However, in his book *Love and Human Separateness* (1987) Ilham Dilman distinguishes between two forms of knowing another person. He writes:

We say we know someone with whom we have worked, someone by whose side we have fought. In this sense we know a friend, a comrade, a colleague, a neighbour; a husband knows his wife and a wife her husband. Here 'I know him' means more than 'I know what he is like', though it includes that. In an important sense of 'know' if I know him there are certain things I can ask of him, certain things I can say to him which I cannot ask of or say to a stranger. (Dilman 1987, p. 121)

On the one hand knowing a person can mean that we have information about certain things concerning the person, i.e. we "know what he is like". However, Dilman claims that there is also another form of knowing a person that is integral to our standing in a personal *relationship*.

According to proponents of theory of mind, there is basically only one form of understanding or knowing another person, that is, to "know what the other is like". It is with this conception of understanding as a starting point that the theorists try to explain what it means that we are social beings. This can, in other words, be described as an epistemological conception of understanding. Dilman suggests, however, that the other form of knowing is of more fundamental importance if we want to understand what it means that we are social beings and what it means

to understand another person. He argues that the concepts of knowledge and understanding in an important sense gain their meaning through our ways of being involved with each other, through talking and doing things together, and then also through sharing close relationships.

The epistemological conception of understanding that can be discerned in theory-of-mind theories is reflected in several more specific forms of explaining interpersonal understanding. These are; a tendency to consider a third-person perspective as basic for what it means to understand other people, a tendency to consider interpersonal understanding as based on an analogical mechanism of imagination, and a tendency to consider the human being in body-mind dualistic terms. I shall briefly describe these three patterns of ideas.

As I maintained above, one central assumption of theory-of-mind theories is that interpersonal understanding consists in a third-person perspective. This largely forms how interpersonal understanding is described. For instance, according to Alvin I. Goldman (2006, p. 3), a central social feature of human life is that we attribute mental states to self and others. Jane Heal (1995) writes “We frequently make judgements about the thoughts and feelings of those about us. We use those judgements in explanations and as the basis for predictions of their future thoughts and actions.” (Heal 1995, p. 33) Statements like these are very common within theory-of-mind theories. The assumption is that interpersonal understanding can be described as a matter of *predicting, judging, explaining, describing, recognizing, making sense of*

and *attributing mental states* to other *minds*. Indeed we do often predict, explain and try to make sense of other people's behaviour. The problem arises when philosophers or psychologists claim that this is the *one and only* or the *most basic* or the *most important* way in which we understand other human beings. That is, there is a tendency to describe interpersonal understanding as if it generally and basically consisted in a *third-person perspective* where we *think about* the other person. The suggestion in theory-of-mind theories is then not merely that we quite often think about other people and wonder what they are up to. The suggestion is that such a perspective is *basic* for what it means to understand another person. This assumption is not often made explicit but is something that can be seen in the form of the arguments and in the form of the examples that are invoked in the theories. It is also an idea that can be seen in the way certain empirical observations are made and in the way experimental situations are constructed and results are described.

Dilman claims, however, in the earlier quote, that it is in our mutual *engagement* with each other that questions about knowledge and understanding have meaning. This also means that it is a central part of what it means to understand another person that we can share a *close personal relationship* with him or her, that we can share a *long life history* with the other person, and that we can be engaged in *conversations* with each other; conversations that often get their character and meaning from the long shared life history.

The emphasis on thinking has also been criticised both inside and outside the branch of theory-of-mind theory. Empathy theorists have been critical of a too rationalistic and individualistic conception of human life. As I have noted earlier, empathy theorists consider the fact that we often respond *spontaneously* and *emotionally* to other people as an important expression of understanding. However, even if empathy theorists have an important hunch here about the spontaneous and emotional character of interpersonal understanding, their way of explaining this in analogical terms becomes problematic. Even if a third-person perspective is not as explicit, the analogical model of explanation that empathy theorists work with is also ultimately based on a third-person conception of interpersonal understanding. It is a central assumption in theories on empathy that I *observe* others and *then* I feel the same as the other and *then* I understand how the other feels.

However, the argument from analogy is not only problematic because it assumes that we generally understand other people by observing them (i.e. a third-person perspective) rather than in some sense engaging with the other person. It is also problematic because it assumes that we can only understand other people *indirectly*, while it is assumed that we have *privileged access* to our own thoughts and feelings. According to Norman Malcolm (1972), this is a central assumption in empiricist philosophy.

[...] for philosophers of an empiricist inclination it has seemed a matter of course that we learn from

introspection what thinking, remembering, and perceiving are. [...] The ostensive definitions are 'private'; they take place in each one's mind, no one else can be aware of, can directly know, those phenomena in my mind that I name 'thinking,' 'remembering,' or 'believing'. (Malcolm 1972, pp. 15-16)

This assumption, that we understand the meaning of "mental" concepts by introspection, leads to the idea that we only understand other people indirectly. The argument from analogy then comes to seem like a good explanation of interpersonal understanding. The idea is that interpersonal understanding is dependent on an introspective first-person perspective which enables us to understand other people through analogical imagination.

The above described tendencies are also connected with a tendency to talk of the human body as if it was a surface that we observe. The result is that a body-mind dualism is often implicit in theory-of-mind theories. The assumption is that by observing the other person's physical movements I can infer what goes on in his mind. Generally theory-of-mind theorists are not expressly body-mind dualists or solipsists. They do think we can understand other people's bodily expressions, and they do think human beings are essentially social beings, and this is also their main interest. Many of the theory-of-mind theorists would probably be critical of classical body-mind dualism. But nevertheless I will claim that their way of explaining interpersonal understanding often rest on dualistic assumptions.

There are several reasons why theory-of-mind theorists tend to talk of the human body as a surface. One reason is that if interpersonal understanding is taken to consist in a third-person perspective, the human body also becomes something that we describe as something we *observe*. From such a point of view a surface perspective on other people's bodies easily becomes appealing. Another, though related, reason why theory-of-mind theorists tend to talk of the body as a surface has to do with the idea that we understand other people through a mechanism or method of analogical imagination or analogical emotional reactions. If one thinks that interpersonal understanding takes an analogical form it becomes important that human beings are *similar*. And then it also becomes important that we have physically similar bodies. From this it easily follows that one adopts a kind of physiological-anatomical perspective on the human body. A third reason why theory-of-mind theorists tend to talk of the body as a surface has to do with the fact that they are influenced by psychology and psychology has a long history of influence from physiology. René Descartes is generally referred to as the father of body-mind dualistic thinking. However, Descartes was strongly influenced by modern medical science and physiology. In a somewhat similar sense physiology has had a strong influence on psychology. Much of the body-mind dualistic thinking that can be found in theory-of-mind theory is not in a clear sense influenced by Cartesian philosophy but rather by physiological and anatomical ways of speaking of the body that has also largely influenced psychology. Important to

see here is then that the “surface” perspective on the human body gets support from several aspects of theory-of-mind theories. First, the tendency within theory-of-mind theories to assume that interpersonal understanding is epistemological in kind makes the researchers inclined to assume that interpersonal understanding is based on a third-person perspective. Second, the assumption that interpersonal understanding is based on a third-person perspective makes the researchers inclined to talk of the human body as something we observe from a distanced viewpoint. Third, the fact that theory-of-mind theories are often based on the argument from analogy makes the researchers inclined to consider human bodies *similar*, which makes them pay attention to the similarity of bodily expressions. Fourth, theory-of-mind theories are closely linked with empirical research in psychology, which again is influenced by physiology. This makes the researchers even more inclined to stick to a physiological-anatomical way of speaking of the body.

To conclude so far, I have above tried to point at some general problems of theory-of-mind theories. The main problems can be said to be that theory-of-mind theories are based on third-person, analogical, and also body-mind dualistic perspectives that are all reflective of an epistemological conception of interpersonal understanding.

The relationship between theory-of-mind theories and empirical research

An important aspect of theory-of-mind theories is that they are deeply embedded in empirical research. A central aim of the thesis is therefore to reflect on the relationship between philosophical theory and psychological empirical research. One can say that there are two kinds of empirical source that theory-of-mind theories are based on. On the one hand many of our everyday experiences often seem to support the various theory-of-mind theories. On the other hand theory-of-mind theories are often connected with various kinds of experimental research. In this sense theory-of-mind theory can often appear to rest on a common-sense conception of human, and animal, life as well as on scientific findings.

For instance, Andrew Meltzoff's research on infants' imitation of facial expressions is not merely reflective of a certain kind of philosophical theory that correlates with certain results in experimental research. Meltzoff's research also reflects the fact that the human face is of great importance in our life with each other. There is also something fundamentally important in the facial contact we can have with a child. In this sense Meltzoff's studies on imitation reflect the importance of the human face in real life. Further one might also note that imitation is connected with learning on a larger scale. Children learn from others and much of this learning does take the form of imitation.

The suggestion that there is an empirically observable link between empathic imagination and compassion can also appear to be reflected in our everyday experiences. Often we are emotionally moved by the sight of another person's suffering. We can find it unbearable to look at an injured person. It is as if *we feel* the other person's suffering. And often we also imagine other people's suffering, we worry etc. Emotional reactions as well as imagination are in this sense often a central part of compassion. Often it is also the case that if we have experienced something similar to another person we also feel we understand how the other person feels. A common way of talking about compassion is that we say that we feel *with* the other person. Sometimes we might also say that we *identify* with the other. In the many versions of the Golden Rule there also seems to be a form of reasoning that is based on our capacity to imagine in an analogical sense: "Love thy neighbour as thyself". Sometimes we use arguments from analogy when we try to make a child reflect on his own actions towards others; "Imagine how would *you* feel if you were left alone like that by your friends?!". These ways of talking and these ways of responding emotionally to other people's suffering seem to support the suggestion that compassion is based on an analogical method of imagination, a form of explanation that in some way or other often comes up in theory-of-mind theories.

When it comes to the more rationalistic theory-of-mind theories there seem also to be an abundance of cases in real life that support these theories. We do often reflect on other people and wonder what a person might think. As I

have already mentioned, the theory-of-mind research on autism is also closely linked with the fact that children with autism do have severe social deficits that often have to do with inabilities to understand other people. Evolutionary theories that are connected with both rationalistic theories of theory of mind as well as with emotivistic theories can also appear commonsensical. We are all aware of the fact that many animals (including human beings) are social by nature. There are also many interesting similarities in how certain animals, such as apes, behave socially and how humans can behave. There are many cases where animals clearly respond with care towards others. In this sense the psychological empirical research on theory-of-mind theory gain support from our everyday observations of human and animal life.

However, not only are theory-of-mind theories supported by a large number of observations in real life. As I have suggested, they also claim support from a large amount of experimental research. However, even though theories on theory of mind largely lean on empirical observations, I will argue that the empirical observations are often *formed* by the earlier described patterns of thinking such as a tendency to emphasise a third-person perspective of understanding, a tendency to consider interpersonal understanding to be based on analogical imagination and a tendency to talk of the human being in body-mind dualistic terms. If a researcher has a confused conception of interpersonal understanding the confusion will often not be solved by empirical research but merely enhanced. My intention is not to criticise psychological

empirical research in general. I also agree with theory-of-mind theorists that the empirical results in the research can be interesting and that we may learn something from the empirical research. My aim is to criticise how the results of the empirical research are interpreted, as well as to criticise how the design of the experiments are formed by problematic philosophical assumptions.

The fact that empirical research *can be* problematic does not mean that empirical research *always must be* problematic. My critical discussion of empirical research is, I think, in line with what Peter Winch says. In *The Idea of a Social Science and its Relation to Philosophy* (1958) he writes:

Philosophy, [...] has no business to be anti-scientific: if it tries to be so it will succeed only in making itself look ridiculous. Such attacks are as distasteful and undignified as they are useless and unphilosophical. But equally, and for the same reasons, philosophy must be on its guard against the extra-scientific pretensions of science. (Winch 1958, p. 2)

Winch further argues that there is often a tendency to mistake philosophical questions for empirical ones. That is, it can be the case that one has a confused understanding of one's original question. This confusion can make one inclined to think that the question can be solved by empirical observation. The point here is *not* that philosophers never can learn anything from empirical research or from observations of real life. Empirical research can be of great importance also for a philosopher, but if the original question is obscure the empirical

connection will also easily become so. In order for us to understand what we should do empirically we must know what our question means. And sometimes when we come to understand the character of our question we might also realise that the question cannot be answered by empirical observations.

Another reason why empirical research can be problematic is that it may be too narrow in focus. This narrow focus can, again, be due to the confused character of our original question. In his *Philosophical Investigations* ([1953] 2001) Ludwig Wittgenstein writes: “A main cause of philosophical disease—an unbalanced diet: one nourishes one’s thinking with only one kind of example.” (Wittgenstein [1953] 2001, §593) Even if we make careful and detailed observations of a phenomenon, we might still have a too one-sided or too restricted focus. The reason we have a too restricted focus can, again, be because we are influenced by a certain philosophical theory.

Theory-of-mind theorists often claim that they describe certain *underlying* cognitive features of interpersonal understanding. It is then also often assumed that by creating a restricted research environment we will be able to study such general cognitive functions. These assumptions are often reflected in a predilection for experimental research methods. Experimental research can be appealing because of the appearance of clarity in the results. The appearance of clarity in the results creates the impression that a general underlying cognitive mechanism can be discerned. At the same time, experimental research may be shaped by the research question in ways that affect

the results in problematic ways. One of my suggestions will be that proponents of theory of mind often tend to accept experimental research methods (and results) too readily. The reason for this is, however, not sloppiness, nor is the reason merely that the researcher is influenced by methods used in natural science (though I do think natural scientific research methods have had a great and problematic influence on certain fields of psychology). Rather there is often a certain *form* of experimental research that fits well with the conception of understanding that theory-of-mind theories rest on. Much of the experimental research within theory-of-mind theory is constructed in a kind of one-directional manner. The subject observes something or someone and then reacts to what he observes, or describes what he observes. That is, the situations are constructed as subject-object situations. Often the situations are also constructed in a passive sense where the subject should *observe* something rather than *do* something. This way of constructing experimental situations has been criticized by Urie Bronfenbrenner (1978). According to Bronfenbrenner, it is generally accepted among psychological researchers that human beings interact with each other and that development of understanding is integral to such interaction. However, even though this is generally acknowledged, it is seldom taken into account in empirical research practice. Instead researchers tend to construct experimental test situations that consistently take a one-directional form. Bronfenbrenner suggests that the one-directional form of psychological experiments derives from the influence from physics. For a long time physics

provided the ideal model for what it means to conduct objective scientific research.

However, when it comes to theory-of-mind theories I claim that there is also another reason. The one-directional and passive form of the experiments are appealing if one has an epistemological (and often empiricist, though not necessarily⁸) perspective on interpersonal understanding. From such a perspective understanding *is* one-directional in its character. It is the subject who observes the object, even if the object happens to be another person. What is ignored then is how our understanding of each other has its meaning as part of our reciprocal responsiveness and as part of our personal relationships. The whole philosophical concept of other minds rests on such a one-directional (non-relational, non-engaged) conception of interpersonal understanding. The problem with the one-directional form

⁸ It is not only researchers of an empiricist kind that have been influenced by one-directional experimental research practice. Somewhat similarly one-directional experimental studies have also been made concerning logical understanding. Jean Piaget (1971) (1974) tested children's development of logical reasoning. Another researcher that can be mentioned here is the sociologist A.R. Luria (1976) who studied illiterate people in remote areas of the Soviet Union. Both Piaget and Luria wanted to study the development of logical reasoning. I will not discuss Piaget or Luria at length here, or their conception of logical understanding. I merely want to note that it is not only empiricist philosophers or psychologists who are inclined to take a one-directional empirical research approach. In *Experience and the Growth of Understanding* (1978) D.W. Hamlyn discusses Piaget's conception of the development of understanding. In *In Defence of Informal Logic* (2000) Don Levi discusses Luria's empirical research.

of experimental research methods then, when it comes to theory-of-mind theories, is not only that the research can tend to be too restricted or narrow or that the research is too highly influenced by natural scientific research methods, but that the narrowness of the empirical research manifests a philosophical tendency to think of interpersonal understanding in third-person, analogical or body-mind dualistic terms, which are all in some sense epistemological perspectives on understanding.

Does this mean that the problematic relationship between theory-of-mind theories and empirical research would be solved if the researchers simply stopped making experiments and instead were to observe “real life”? Not necessarily. As I have noted earlier, much of the empirical research that is connected with theory-of-mind theories is in some way or other inspired by certain observations of real life. It would be a mistake to say that the main problem with experimental research that is built on theory-of-mind theory is that it is *not* concerned with real life. On the contrary, the empirical research often reflects our life in some way or other and the reason the researchers begin to construct experiments and to theorise is often the fact that they are intrigued by certain aspects of real life. It is then important to acknowledge that the patterns of thinking within theory-of-mind theories do not merely arise out of classical (problematic) philosophical assumptions (such as body-mind dualism, the argument from analogy or a third-person perspective). Nor do they merely arise from tendencies to construct reductive empirical experiments on the basis of philosophical theories. On the contrary, theory-

of-mind theories often mirror our life in one way or other, and often the empirical research does so too. This is a further reason why the theories in question can appear so convincing.

Still, even if theory-of-mind theories reflect certain aspects of real life, it can also be the case that real life observations lead a researcher astray. One reason for this is that even when we observe real life situations we are often inclined to assume a reductive perspective. One such way is by merely sticking to what Wittgenstein calls a “one-sided diet of examples”. In his article “Trying to Keep Philosophy Honest” (2005) Lars Hertzberg suggests, in line with Wittgenstein, that when we try to think of examples we often think merely of examples that fit well with our theory, or with our idea of how things are, and we ignore a large part of examples that might go against our thinking. This problem is, of course, not very easy to avoid since it is not always so easy to really see what examples actually are relevant or not for one’s question.

Another way a reductive perspective can form our observations of real life is when we start to use a theoretical vocabulary to describe ordinary life situations. Hertzberg writes: “We test words on our tongue in the solitude of our study, and in doing so we grossly underestimate our inability to imagine the real life of the expressions we are considering.” (Hertzberg 2005, p. 81) Often the problem is also that a certain kind of professional theoretical vocabulary quickly gets accepted among philosophical or psychological theorists. In this sense the problem with how we get blinded by the use of certain

words is not merely an individual problem but largely a cultural problem in scientific discussions. This may concern words such as “altruism”, “empathy”, “theory of mind”, “mind”, “intention”, “belief”, “emotion”, “imagination”, etc. Partly the problem is that researchers invent new scientific words that can have an unclear, reductive or too generalising meaning. This is, for instance, the case with words like “altruism”, “empathy” and “theory of mind”.

The problem is also that words we use in ordinary life are given a *different* meaning in the philosophical and psychological discussions without the researchers noticing it themselves. This is, for instance, often the case with the words “belief”, “intention”, “knowledge” and “human body”. Within theory-of-mind theories these words are often used in a highly specialised or reductive sense though the theorists might not acknowledge it themselves. The field of theory-of-mind theory contains many such uses of expressions that are largely accepted even though the meaning of the expressions actually may be unclear. In this sense we are by no means safe from “extra-scientific pretensions” even if we look at real life. And, as Hertzberg notes, even if we use “ordinary words” we may still have lost the sense for how these words actually are used in ordinary life.

I already noted that theory-of-mind theorists often assume that they describe underlying or general cognitive functions of understanding. One can then get the impression that a reductive, or in other ways specialised use of vocabulary, as well as restricted empirical research,

should be all right. My criticism of theory-of-mind theories often takes the form of suggesting that the use of certain concepts is unclear, as well as suggesting other ways of interpreting experimental results, or describing broader patterns of interpersonal relationships and interpersonal responsiveness than theory-of-mind theorists describe. One can get the impression that my descriptions of human social life are on a “social surface level” while the theory-of-mind theorists’ descriptions are on a “deep cognitive level”. However, my aim is not to contrast a “deep cognitive level” with a “social surface level” of human life but to provide alternative descriptions of interpersonal responsiveness, and thereby also to point to certain conceptual confusions within theory-of-mind theories that are reflected in the empirical research. The aim is thereby to dissolve the impression that theory-of-mind theories describe a basic underlying cognitive level of interpersonal understanding.

To conclude then, a central aim of the thesis is to discuss the relation between theory-of-mind theories and their relation to empirical research as well as their relation to real life observations. I will claim that certain presuppositions often shape the research as well as the real life observations so that certain results are reached or so that the real life situations seem to fit well with a theoretical perspective, without the researchers acknowledging it themselves. Although I will critically discuss empirical research my aim is not to contribute with new empirical research. My aim is to point at certain conceptual confusions in theory-of-mind theories and also

then conceptual confusions that are reflected in how the empirical research is built up and results are interpreted.

Summaries of the chapters

Chapter 1: Rationalistic and emotivistic theories of altruism

In this first chapter I contrast emotivistic and rationalistic theory-of-mind theories in the context of evolutionary psychology and specifically concerning altruism. I begin by describing Cosmides and Tooby's theory of altruism. According to them, altruism consists of an algorithmic reasoning mechanism by which we calculate how to act in order to maximize fitness when dealing with others. I contrast their theory with de Waal's theory on altruism as consisting of a mental mechanism of empathic imagination. Cosmides and Tooby's as well as de Waal's perspectives are expressive of two problems. First, the reliance on a transactional model of interpersonal relationships. Second, the view of interpersonal understanding as consisting of an ability for analogical imagination. By reflecting on various examples of the natural and social form of our life I question the above mentioned perspectives and point to alternative ways of understanding human social life and the ways we can care for each other. In this context I also bring in Ludwig Wittgenstein's reflections on primitive reactions and language.

Chapter 2: Theory of mind and infants' imitation of facial expressions

In this second chapter I focus on the conception of the human body that can mainly be seen in emotivistic theories. This emotivistic conception is also often called "simulation theory". According to simulation theory interpersonal understanding is dependent on a capacity for analogical imagination which again is dependent on the capacity to simulate bodily reactions. The chapter deals with certain psychological experiments on imitation that have been made with infants. The psychologists Andrew Meltzoff and M. Keith Moore have shown that newborn infants have a capacity to imitate other people's facial expressions. These findings suggest that children have a theory-of-mind function by which they learn to decipher other people's facial expressions and thereby also eventually to understand other people's intentions. One of my aims in this chapter is to discuss how these empirical observations should be understood. My aim is not to deny the empirical findings per se. Rather I argue that certain philosophical assumptions shape the research methods so that certain results are reached. I also argue that the appearance of clarity in the experimental results is due to the strongly restricted focus.

Chapter 3: Autism and theory of mind

The theme of this chapter is the syndrome of autism and research suggesting that persons with autism lack a theory-

of-mind function, i.e. that autism can be described as “mindblindness”.

Autism is a pervasive developmental disorder that affects a person’s life extensively. The most striking and central feature in autism is the lack of social responsiveness. A child with autism is largely unresponsive to other people, and often has serious deficits in language. Such a child generally does not seem to acknowledge the presence of other people, does not respond to other people’s talk, and generally does not play with other children etc. Such traits in autism can appear to point towards an inability to see that other people have minds.

In this chapter I discuss the theory that autism consists in “mindblindness”. I reflect on several empirical studies that seem to imply that children with autism lack a theory-of-mind function. Even though the empirical findings are important and even though they do seem to point to some form of severe social dysfunction, I claim that the findings are more ambiguous than the theory-of-mind theorists acknowledge. I also claim that the image of autism as mindblindness is created by a one-sided conception of ordinary forms of interpersonal understanding as well as a one-sided and restricted focus when discussing the responses of children with autism. By looking at Clara Claiborne Park’s (1967) (2001) biographical books about her autistic daughter I try to bring in a more varied picture of autism. I do not want to deny that persons with autism have various severe problems with interpersonal understanding. However, I maintain that these problems cannot be understood from the point of view that ordinary

interpersonal understanding consists in a theory-of-mind function.

My intention in this chapter is twofold. On a specific level the aim is to discuss and question the claim that persons with autism lack a theory-of-mind function. However, on a broader level, which connects it with the other chapters in this thesis, my aim is to once again discuss the relationship between empirical research and theory-of-mind theory.

Chapter 4: The relationship between empathic imagination and compassion

In this last chapter I discuss certain theories suggesting that there is a relationship between empathic imagination and compassion. I begin by discussing certain experimental investigations on empathy that are considered to prove that there is a psychological causal link between empathic imagination and compassion. I discuss the results of the experiments and argue that they are more ambiguous than might seem at first glance. I question the experiments on two accounts. On the one hand, the experimental context is so strongly reduced that it becomes unclear in what sense the participants can be said to understand that another person is suffering. Further, I maintain that the instructions that the experimenters give to the test subjects are leading. This creates a further ambiguity in how the responses should be understood.

By discussing some biographical descriptions of suffering as well as of compassion I claim that our failures

to understand another person's suffering are of a quite different character than what is assumed in the cognitively oriented theories on empathy. At the same time, however, there are some features in our ways of reacting to another person's suffering that can partly explain why a cognitive capacity for analogical imagination can be appealing as a philosophical theory. Also both mind-body dualistic as well as solipsistic forms of thinking gain credibility when one reflects on suffering. However, even if there are traits in suffering that can give a certain comprehensibility to philosophical perspectives such as empathy, body-mind dualism or solipsism, this still does not make these perspectives unproblematic.

In the last part of the chapter I discuss the idea that empathic imagination can be used for both good and evil purposes. I argue that knowledge is not necessarily always expressive of understanding. The assumption that there is a cognitive method of empathic imagination that we can use for good or evil purposes is based on a conception of knowledge as something neutral, something that is independent of our attitudes and of our responsibility towards other people. I argue that this is a confusion regarding the concept of knowledge which is also reflected in the idea that there is some such thing as a cognitive method of empathic imagination.

Chapter 1: Rationalistic and emotivistic theories of altruism⁹

1.1 Introduction

In this chapter I discuss two perspectives on altruism. I do so by looking at some central modern evolutionary theorists, connecting these with two traditions of philosophical thinking. The modern evolutionary theorists I discuss are the psychologist Leda Cosmides, the anthropologist John Tooby and the primatologist and ethologist Frans de Waal. These represent two different theoretical perspectives on altruism. Cosmides and Tooby can be said to have a rationalistic perspective on altruism. According to them, altruism consists of an algorithmic reasoning mechanism by which we calculate how to act in order to maximize fitness when dealing with others. According to de Waal, on the other hand, altruism is based on empathic imagination. These two modern perspectives on altruism are linked with two theoretical traditions. Cosmides and Tooby's rationalistic perspective can be traced back to Thomas Hobbes, Herbert Spencer, William D. Hamilton and John Maynard Smith. De Waal's emotionally oriented "empathic" perspective on altruism can, in turn, be traced back to Adam Smith, Charles

⁹ An earlier version of this chapter has been published in Gustafsson (2012).

Darwin and Edward Westermarck. Darwin and Westermarck can be regarded as important contributors to both of these two traditions of thought, although their main approaches stand closer to de Waal's.

The main problems about the approaches I discuss are the reliance on a transactional model of interpersonal relationships and the view that interpersonal understanding consists in an ability for analogical imagination. I will also maintain that the very concept of altruism may obscure more than clarify the meaning of care and compassion in human life. Finally I also claim that the two perspectives on altruism are, in some important ways, not actually opposites but share similar, in part questionable, presuppositions about the character of human life.

Even if both the above mentioned perspectives are problematic, many observations and descriptions made by these researchers on how people (and animals) sometimes behave and reason in real life can be made out to support their views. However, even though the two perspectives on altruism offer interesting and also insightful observations of human life, the researchers tend to give problematic explanations for their observations. I suggest that some of their descriptions are too restricted in focus or too one-sided or generalising. The reductive and generalising descriptions of human life are expressive of how the researchers are influenced by classical, but problematic, conceptions of our shared human life and interpersonal understanding. By bringing in alternative examples and alternative descriptions of the behaviour

they cite for support, I try to point to other ways of describing human life. In this context I also bring in Ludwig Wittgenstein's reflections on primitive reactions and language.

Still, even if I question the two perspectives on altruism I also think that especially the "theorists of empathy", i.e. Adam Smith, Darwin, Westermarck and de Waal offer many sensitive reflections on the natural character of our shared human life; reflections that also in many ways resemble Wittgenstein's reflections on the natural form of human life.

1.2 Altruism as transaction

According to Cosmides and Tooby, the human mind has adapted over thousands of years, so that we have become the social and moral beings of today. Through evolutionary history, we have developed a mental skill of calculating how to act optimally in our dealings with others. All such dealings are formed so that they on the whole tend to enhance fitness i.e. the propagation of copies of the agent's alleles in subsequent generations.¹⁰ Cosmides and Tooby (2008) describe this view as a social contract theory.

¹⁰ In *The Adapted Mind* (1992) Cosmides and Tooby distinguish between two meanings when they talk about "fitness". They write: "Previously, evolutionary biologists spoke of a design's 'Darwinian fitness': its effect on the number of offspring produced by an individual who has the design. But since Hamilton, one speaks of a design's 'inclusive fitness': its effect on the number of offspring produced by an individual who has the design *plus* its effects on the number of offspring produced by others who may have the same design—that individual's relatives—with each

Social contract theory is based on the hypothesis that the human mind was designed by evolution to reliably develop a cognitive adaptation specialized for reasoning about social exchange. [...] From an evolutionary point of view, the design of programs causing social behavior is constrained by the behavior of other agents. More precisely, it is constrained by the design of the behavior-regulating programs in other agents and the fitness consequences that result from the social interactions these programs cause. (Cosmides and Tooby 2008, pp. 69-70)

Cosmides and Tooby further describe the fitness enhancing systems of reasoning to which our human minds have adapted as “evolutionarily stable strategies” or “ESS”. According to them, human social engagement has evolved because it has proved to be an evolutionarily stable strategy. If our social engagement with others did not enhance fitness it could not have survived for thousands of years.

An evolutionarily stable strategy (ESS) is a strategy (a decision rule) that can persist in a population because it produces fitness outcomes greater than or equal to alternative strategies.¹¹ The rules of reasoning and decision making that guide social exchange in humans would not exist unless they had outcompeted

effect discounted by the appropriate measures of relatedness [...]” (Cosmides and Tooby 1992, p. 168) When Cosmides and Tooby talk about fitness they mean “inclusive fitness”.

¹¹ Cosmides and Tooby refer here to J. M. Smith (1982).

alternatives, so we should expect that they implement an ESS. (Cosmides and Tooby 2008, p. 70)

In another article Tooby, Cosmides, Aaron Sell, Debra Liebermann and Daniel Sznycer (2008) conclude that altruism is the outcome of a nonconscious mental mechanism whereby we calculate how to cooperate with others. They note that human beings tend to be more altruistic towards their family than towards other people. This they define as “kin selection”. Kin selection is, according to them, an evolutionarily stable strategy that enhances fitness. Tooby et al. also speak in this context of a “welfare trade-off ratio” or “WTR”. By a “welfare trade-off ratio” they mean a mental mechanism or a “variable” that regulates how much we ought to help others in order to gain certain benefits. This variable can be “upregulated” when we are dealing with genetic relatives.

[N]atural selection should have designed the human motivational architecture to embody programs determining how high one’s welfare trade-off ratio toward other individuals should be set. [...] kin selection theory tells us that, all else equal, WTR should be upregulated for close genetic relatives, motivating us to help kin more and harm them less than we otherwise would. (Tooby et al. 2008, p. 260)

The degree to which we make sacrifices for others is dependent on a “kinship index” or “kin selection”. Important to note here is also that Tooby et al. think of these systems of reasoning as occurring nonconsciously.

[T]he welfare trade-off ratio, WTR is an internal regulatory variable expressing how much you value j's welfare relative to your own. Its value is nonconsciously expressed in many decisions you make throughout the day—how much chocolate you leave for j, how loud you play your music when j is trying to work, whether to clean up the mess or leave it for j, whether to call home to let j know you will be late. It is computed by a system, the welfare trade-off ratio estimator, that takes into account a specific array of relevant variables [...]. (Tooby et al. 2008, p. 261)

Tooby et al. follow a pattern of thinking that can be traced back to Thomas Hobbes' idea that human social life can be described in terms of a trade-off scenario within the frame of a social contract.¹² However, for Hobbes the social contract was a theoretical explanation of the origins of society. When he described the state of nature as a state of war between individuals he did not claim that this is how things have *actually* originally been. Rather he wanted to paint a *theoretical picture* in order to advance his ideas about how to achieve a peaceful society.¹³ When Hobbes talks of the *natural state* of man as well as of *the social contract* he constructs a theoretical image that is not meant to be taken as a historical or biological hypothesis. The suggestion that what is described is really a biological fact about our species and real mental functions comes later, with the

¹² See, for instance, Hobbes, ([1651] 1996), *Leviathan*, chapters 13–15.

¹³ Another philosopher who, similarly, paints a theoretical picture of human life as originating from a state of nature is Jean-Jacques Rousseau.

evolutionary biologists William D. Hamilton (1964) and John Maynard Smith (1982). Hamilton argues that an organism's survival ought to be understood according to "inclusive fitness". By inclusive fitness Hamilton means that an organism's genes have higher chances to survive if the organism cooperates with relatives or with other organisms of the same species. According to Hamilton, this explains why some species are altruistic. He writes:

The social behavior of a species evolves in such a way that in each distinct behaviour-evoking situation the individual will seem to value his neighbour's fitness against his own according to the coefficients of relationship appropriate to that situation. (Hamilton 1964, p. 19)

Following Hamilton's ideas on inclusive fitness, John Maynard Smith introduced the idea of "evolutionary stable strategies" and argued that "kin selection" and "group selection" are such basic evolutionarily stable strategies. J. M. Smith (1982) also introduced the economic principles of game theory in his reflections on evolution. Cosmides and Tooby largely follow this line of thinking.

1.3 Emotivistic conceptions of altruism

Frans de Waal's perspective on social life and on altruism differs from Cosmides and Tooby's perspective. De Waal is sceptical of a transactional perspective where we fulfil other people's wishes in order to gain advantages. He sees ape and human life as a shared life where we often

naturally and spontaneously help each other *without* wanting anything in return. In *The Age of Empathy* (2009) he has, for instance, the following description of chimpanzees helping each other.

There is in fact so much evidence for altruism in apes that I will pick just a handful of stories to drive home my point [...] At our primate center, we have an old female, Peony, who spends her days with other chimpanzees in a large outdoor enclosure. On bad days, when her arthritis is flaring up, she has great trouble walking and climbing. But other females help her out. For example, Peony is huffing and puffing to get up into the climbing frame in which several apes have gathered for a grooming session. An unrelated younger female moves behind her, places both hands on her ample behind, and pushes her up with quite a bit of effort, until Peony has joined the rest. (de Waal 2009, p. 105)

De Waal has many descriptions like this one of how chimpanzees spontaneously help each other without requiring anything in return. He sees altruism as spontaneous emotional responsiveness that evolves naturally as we grow up with others. In *Primates and Philosophers, How Morality Evolved* (2006) he writes: "Since expressions of sympathy emerge at an early age in virtually every member of our species, they are as natural as the first step" (de Waal 2006, p. 28). The fact that others care for us, according to de Waal, is a natural aspect of how

we grow up and thus also develop emotional responsiveness to each other.

De Waal's perspective can be seen as linked with a different tradition than Cosmides and Tooby's. De Waal follows a tradition of thought that can be traced back to Adam Smith, Charles Darwin and Edward Westermarck. According to Adam Smith, human beings have a natural inclination to form social bonds and relationships and to feel compassion for others. In *The Theory of Moral Sentiments* ([1759] 2002) Smith writes:

How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it. Of this kind is pity or compassion, the emotion which we feel for the misery of others, when we either see it, or are made to conceive it in a very lively manner. That we often derive sorrow from the sorrows of others, is a matter of fact too obvious to require any instances to prove it; for this sentiment, like all the other original passions of human nature, is by no means confined to the virtuous and humane, though they perhaps may feel it with the most exquisite sensibility. The greatest ruffian, the most hardened violator of the laws of society, is not altogether without it. (Smith [1759] 2002, pp. 11-12)

According to Smith, we spontaneously care for others *without* requiring anything in return.¹⁴ Consider again Tooby et al.'s description of such acts as calling home to let one's family know one will be late. Instead of describing this as a trade-off response, one could say that calling home to let one's family know one will be late, is an expression of spontaneous considerateness and care. It is usually not a decision we make on the basis of a calculation, nor is it because we expect something in return that we call home. Very often it is simply something we do because otherwise the family would get worried. In this way we often respond spontaneously with care towards our family (as we of course also often respond spontaneously with irritation or other negative emotions).

Charles Darwin also maintains that human beings naturally have a "social instinct" and an inclination to help each other. However, Darwin differs from Adam Smith in that Darwin's perspective is more biologically oriented. Thus he draws many comparisons between animal and human behaviour. In *The Origin of Species and The Descent of Man* ([1859, 1871] 1872) Darwin writes:

¹⁴ Smith does not specifically use the word "altruism", as there was no such word at the time when he was writing. The French term "*altruisme*" was first introduced into philosophical theory by Auguste Comte in his *Catéchisme Positiviste* (1852). This term was translated as "altruism" in Richard Congreve's English rendering of the text *The Catechism of Positive Religion* (1858). Smith's ideas on compassion are, however, similar to Comte's ideas on altruism, inasmuch as Comte does not think of altruism as being based on self-interest.

[...] the social instincts lead an animal to take pleasure in the society of its fellows, to feel a certain amount of sympathy with them, and to perform various services for them. The services may be of a definite and evidently instinctive nature; or there may be only a wish and readiness, as with most of the higher social animals, to aid their fellows in certain general ways. (Darwin [1859, 1871] 1872 p. 472)

He continues:

The feeling of pleasure from society is probably an extension of the parental or filial affections, since the social instinct seems to be developed by the young remaining for a long time with their parents; and this extension may be attributed in part to habit, but chiefly to natural selection. (Darwin [1859, 1871] 1872, p. 478)

Darwin saw it as a central feature of natural selection that animals have natural “social instincts”. According to him, these instincts originally develop among human beings in the child’s relationship with its parents.

Another philosopher who was influenced by both Adam Smith and Darwin, (and who also has influenced de Waal) is Edward Westermarck. He writes in *Ethical Relativity* (1932):

For my own part I maintain that Adam Smith’s *Theory of Moral Sentiments* is the most important contribution to moral psychology made by any British thinker, and that it is so in the first place on account of the emphasis

it lays on the retributive character of the moral emotions. (Westermarck 1932, p. 71)

Like Smith, Westermarck emphasises our spontaneous responsivity towards each other. And in line with Darwin he claims that the origin of our social instincts can be traced back to certain biological social patterns such as the natural relation between parent and child. Thus, in *The Origin and Development of the Moral Ideas, Volume II* (1917), he writes: "There is one form of the altruistic sentiment which man shares with all mammals and many other animals, namely, maternal affection." (Westermarck 1917, p. 186) He continues:

When the young are born in a state of utter helplessness somebody must take care of them, or the species cannot survive, or, rather, such a species could never have come into existence. The maternal instinct may thus be assumed to owe its origin to the survival of the fittest, to the natural selection of useful spontaneous variations. (Westermarck 1917, p. 188)

To sum up the discussion so far, one can say that there are two main traditions of ideas with regard to altruism that I will discuss. There is the transactional and computational perspective beginning with Hobbes and continuing, with alterations, in Hamilton, J.M. Smith, Cosmides and Tooby. On the other side we have the emotivistic perspective, beginning with Adam Smith and continuing with variations, in Darwin, Westermarck and de Waal. Here our tendency to respond *spontaneously and emotionally* to other

people's plight is emphasised. This perspective also emphasises the fact that we are *born into* certain forms of social relations (mainly families), and that we grow into *habits* of being with and responding to each other.

These two traditions of ideas might appear to be sharply opposed. However, it is not evident that they are so. For instance, when de Waal states that our compassionate responses are not based on any attempt to gain benefits, Cosmides and Tooby might well agree, and reply by saying that such responses of course do not *appear* to involve calculation; it is simply the case that these calculations occur *nonconsciously*. I will return to these questions later.

1.4 Sympathy, empathy and retributive emotions as consisting of analogical imagination

One idea that Adam Smith, Darwin, Westermarck and de Waal have in common is a conception of emotions as reciprocal or (using Westermarck's term) "retributive". This is again connected with a certain conception of the nature of interpersonal understanding. As I already mentioned, all four authors would emphasise our spontaneous inclination to care for others. However, all four also maintain that this inclination can be explained as deriving from a certain function embedded in the mind, even if it is not a computational function.

According to Adam Smith, our care for other people is connected with our capacity to imagine ourselves in the other's situation. This can be seen in his ideas on sympathy.

By “sympathy”, Smith does not mean simply compassion, but any kind of imaginative experience of another person’s feelings that cause us to be emotionally moved by the other. This is, in modern terms, called “empathy”. The explanation for our spontaneous reactions of compassion must be sought in the functioning of our minds and in our capacity to imagine how others feel.

As we have no immediate experience of what other men feel, we can form no idea of the manner in which they are affected, but by conceiving what we ourselves should feel in the like situation. Though our brother is upon the rack, as long as we ourselves are at our ease, our senses will never inform us of what he suffers. They never did, and never can, carry us beyond our own person, and it is by the imagination only that we can form any conception of what are his sensations. Neither can that faculty help us to this any other way, than by representing to us what would be our own, if we were in his case. It is the impressions of our own senses only, not those of his, which our imaginations copy. By the imagination we place ourselves in his situation, we conceive ourselves enduring all the same torments, we enter as it were into his body, and become in some measure the same person with him, and thence form some idea of his sensations, and even feel something which, though weaker in degree, is not altogether unlike them. (Smith [1759] 2002, pp. 11-12)

Smith’s explanation of moral sentiments consists in an argument from analogy. Sympathy is the effect of our capacity to imagine ourselves in the other’s shoes. This

effect can in turn make us feel compassion for the other person's misery.

Darwin and Westermarck argue along similar lines as Smith. Darwin refers to Adam Smith's ideas of sympathy as consisting of analogical imagination. He further suggests that natural selection has increased our tendency to feel sympathy especially for our loved ones.¹⁵

According to Westermarck, morality is the outcome of "retributive emotions":

The moral emotions are retributive emotions. A retributive emotion is a reactive attitude of mind, either hostile or kindly, towards a living being (or something taken for a living being), regarded as a cause of pain or pleasure. (Westermarck 1932, p. 172)

However, he recognizes that the individual's reactions to the causes of his or her *own* pleasure or pain do not yet explain why the individual reacts to the pleasure or pain of others:

Our retributive emotions are, of course, always reactions against pain or pleasure felt by ourselves; this holds good of the moral emotions as well as of anger, revenge, and gratitude. The question to be answered, then, is, Why should we, quite disinterestedly, feel pain calling forth disapproval because our neighbour is hurt, and pleasure calling forth approval because he is benefited? That a certain act causes pleasure or pain to the bystander may be due to the close association that

¹⁵ See Darwin ([1859,1871] 1872) pp. 478-479.

exists between these feelings and their outward expressions. The sight of a happy face tends to produce some degree of pleasure in him who sees it; the sight of the bodily signs of suffering tends to produce a feeling of pain. In either case the feeling of the spectator is due to the fact that the perception of the physical manifestations of the feeling produces the feeling itself on account of the established association between them. (Westermarck 1932, p. 96)

According to Westermarck, when I see another person's bodily expressions this causes me to associate these expressions with certain feelings; and thereby I feel these feelings myself. This analogical mechanism of imagination explains, according to him, why we are inclined to help another person when he is in pain. However, Westermarck also suggests that this spontaneous emotional responsiveness to others develops through habit. In this sense he does not merely consider compassion to derive from a mental capacity of analogical imagination. Rather he claims that our inclination to imagine how other people feel grows out of our habits of shared relationships. It is, according to him, important that we get used to sharing our life with each other, for how our emotional responsiveness also develops.

Frans de Waal's ideas are largely along similar lines as those of Adam Smith, Darwin and Westermarck. As I already suggested, according to de Waal, our inclination to care for each other is grounded in the fact that we are born into social relationships. But he also thinks that it can be explained as consisting of a mental mechanism of

imagination. According to him, both humans and apes have a natural capacity to imagine how others feel; a capacity that is enhanced by our habitual ways of living together.

De Waal, who is more biologically oriented than Adam Smith and Westermarck, also describes empathy as at first originating from a bodily reaction that eventually develops into a full fledged capacity to imagine another's perspective. According to de Waal, then, first we simply have a tendency for what he calls "emotional linkage" or "emotional contagion". Eventually we also develop the capacity to see that others have different perspectives on reality. This is, according to him, part of what it means to feel empathy.

When the emotional state of one individual induces a matching or closely related state in another, we speak of 'emotional contagion' [...] With increasing differentiation between self and other, and an increasing appreciation of the precise circumstances underlying the emotional states of others, emotional contagion develops into empathy. Empathy encompasses—and could not possibly have arisen without—emotional contagion, but it goes beyond it in that it places filters between the other's and one's own state. (de Waal 2006, p. 26)

According to de Waal, this capacity for empathy is *not* something that can be explained as deriving from social competition, but derives from a need for cooperation. Human beings as well as apes have a natural inclination to

become emotionally affected by others because we have a natural need to cooperate. The natural habit of cooperating in turn enhances our inclinations to become emotionally affected by each other.

I am personally convinced that apes take one another's perspective, and that the evolutionary origin of this ability is not to be sought in social competition, even if it is readily applied in this domain [...], but in the need for cooperation. At the core of perspective-taking is emotional linkage between individuals—widespread in social mammals—upon which evolution (or development) builds ever more complex manifestations, including appraisal of another's knowledge and intentions. (de Waal 2006, p. 72)

This “emotional linkage” eventually leads to a more advanced capacity to imagine another's perspective, which de Waal calls empathy.

Even if I think there is much truth in these ideas on how our responsiveness to each other, and our care for each other, grow out of our habits of sharing each other's life, I also think it is problematic to claim that, at the bottom of our responsiveness, there is some such thing as an analogical mechanism of imagination. De Waal describes the empathic mechanism as follows:

[...] at the core of the empathic capacity is a relatively simple mechanism that provides an observer (the 'subject') with access to the emotional state of another (the 'object') through the subject's own neural and bodily representations. When the subject attends to the

object's state, the subject's neural representations of similar states are automatically activated. The closer and more similar subject and object are, the easier it will be for the subject's perception to activate motor and autonomic responses that match the object's (e.g., changes in heart rate, skin conductance, facial expression, body posture). This activation allows the subject to 'get under the skin' of the object, sharing its feelings and needs, which embodiment in turn fosters sympathy, compassion, and helping. (de Waal 2006, pp. 37-38)

I agree with de Waal (and Smith, Darwin and Westermarck) that it is a common feature that we are emotionally affected by others and that a kind of emotional contagion often occurs. But the question is whether this can be understood as expressive of an analogical mental mechanism, as Smith, Westermarck and de Waal describe it. And the question is also whether compassion can be understood in this sense as deriving from analogical imagination.

De Waal also assumes that it is central for our capacity to be emotionally moved by other people that we have physically similar bodies. There is, I think, something partly correct and partly problematic in de Waal's suggestion that the human bodily form is of importance for our capacity to care for other people. I shall continue to discuss this in the next section.

1.5 Criticism of the argument from analogy

As I have noted above, according to Smith, Darwin, Westermarck and de Waal, compassion is based on a cognitive capacity for analogical imagination. This idea reflects certain problematic assumptions. It reflects the assumption that I understand the other person's emotions by first understanding what the emotion means in my own case, and by then projecting these feelings onto the other person. Such an idea rests on a classical form of philosophical scepticism where the idea is that we can only really know what we feel ourselves. In other words, the idea that compassion is based on the use of an analogical method of imagination (i.e. empathy) is expressive of a tendency to assume that the second-person perspective (i.e. how we respond to other people) is dependent on the first-person perspective, when we talk about sensations. The idea is that the second-person perspective on sensations is a kind of copy of the first person perspective. This assumption can be seen in the explanations of how empathy works.

However, we do sometimes imagine how we would feel in another person's situation, and these thoughts can sometimes be connected with compassion, but this is merely one form our care for another person can take. Our care for another person can equally well take the form that we do *not* use our imagination. To take an example, you see a child hit by a car and you run instantly to help the child. The only thing that passes your mind might be something like "Oh my God!" That is, often we simply

react to suffering and no imagination is needed. But sometimes we might also imagine things. For instance, in the evening after the accident you cannot stop thinking about the child. Your thoughts about the child can also be entwined with how you think about your own children. Here one could say that a kind of analogical imagination occurs. However, this analogical imagination is in itself already a kind of moral response, it is one way you are shocked and can't stop worrying. The fact that reactions of this kind can occur does not show that this is a basic cognitive mechanism that enables us to understand other people. We can respond with compassion in many different ways to other people's suffering, but this does not mean that *one* of these ways of responding is the basis for all other kinds of responses. It can also be noted that in the case above the analogical imagination is a reaction *after* the accident. It is, in this case, not something that precedes, and motivates, the helping action.

The problem with Adam Smith, Darwin, Westermarck and de Waal is their conception of how analogical imagination works and of its importance. They assume that it is a general and basic cognitive method or function that enables us to understand other people and that, because we imagine ourselves in the suffering person's place, enables us to care for the other.

Wittgenstein suggests another way to consider what it means to understand another person, and thus also what it means to care for another person's suffering. In *Zettel* ([1967] 1981) he writes:

Only surrounded by certain normal manifestations of life, is there such a thing as an expression of pain. Only surrounded by an even more far-reaching particular manifestation of life, such a thing as the expression of sorrow or affection. And so on. (Wittgenstein [1967] 1981, §534)

A few paragraphs later he continues:

It is a help here to remember that it is a primitive reaction to tend, to treat, the part that hurts when someone else is in pain; and not merely when oneself is—and so to pay attention to other people's pain-behaviour, as one does not pay attention to one's own pain behaviour. (Wittgenstein [1967] 1981, §540)

Partly Wittgenstein's reflections here resemble Adam Smith's, Darwin's, Westermarck's and de Waal's ideas in the emphasis on the spontaneous primitive reaction to care for other people. An important difference, however, is that Wittgenstein does *not* claim that such a reaction can be explained as deriving from a general cognitive capacity for analogical imagination. According to Wittgenstein, it is against the background of a broad shared pattern of life—a life where we naturally share close relationships with others and where we then also respond spontaneously to each other in various ways, by for instance sorrow or affection that the concept of pain has meaning. It is then also as part of such a broad shared pattern of life that our spontaneous responsiveness to each other sometimes can be reflected in how we think of others. Wittgenstein does

not deny that we may sometimes imagine in an analogical sense. But he does not maintain that such a form of imagination is a basic cognitive mechanism that enables us to understand others.

However, as I have already claimed, even if Smith, Darwin, Westermarck and de Waal partly explain altruism in terms of a mental method or function of analogical imagination, they also partly have a perspective that is similar to Wittgenstein's. Wittgenstein says that it is a primitive reaction to tend to another's pain. Smith, Darwin, Westermarck and de Waal say that we spontaneously respond to other people's suffering. Darwin, Westermarck and de Waal also emphasise the importance of habit for how we grow into being responsive to each other. According to Wittgenstein, the expression of pain has its meaning because it is part of a broader pattern of a shared human life.

Does this mean that Smith, Darwin, Westermarck and de Waal actually think along similar lines as Wittgenstein? Even if there are similarities, there are also important differences. As was already suggested, one important difference lies in the cognitive and also essentialistic role that Smith, Darwin, Westermarck and de Waal give to such concepts as imagination and understanding (treating them as epistemic elements that so to speak give compassion a justification in facts). When Wittgenstein says that we can only understand the expression of pain if we see it as part of a broader pattern of a shared life, he suggests that it is problematic to try to give *one* perspective (i.e. the first person perspective) a *basic explanatory* role if we want to

understand such aspects of human life as pain or compassion. It will be equally problematic to give analogical imagination such a basic explanatory role. The inclination to think that interpersonal understanding rests on a cognitive function of analogical imagination, or that it rests on analogical bodily responses, is reflective of an inclination to ignore that we are involved in each other's life in a multitude of ways. The inclination to use generalising expressions for our care for other people, such as the expression "altruism", is linked with the above tendencies.

There also appears to be a similarity between the ways de Waal and Wittgenstein talk of the human bodily form. According to de Waal, it is an important aspect of empathy that we have physically similar bodies. This is connected with how he thinks of interpersonal understanding as dependent on analogical bodily responses. Again, if one looks at some of Wittgenstein's remarks it can appear as if he were saying something similar as de Waal. In *Philosophical Investigations* ([1953] 2001) he writes:

[...] only of a living human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious. (Wittgenstein [1953] 2001, §281)

One can take this passage to mean that in order for us to understand other people they must be physically similar. However, David Cockburn (1985) suggests another way to understand Wittgenstein's remark. He writes:

[...] the possibility of thinking of another as being in pain, afraid, angry and so on is tied to the possibility of feeling certain things towards them—of responding to them in certain ways. And the possibility of responding in the relevant ways is tied to their having a certain form. (Cockburn 1985, p. 491)

In another paper (1994) Cockburn writes:

We turn away in horror from the sight of someone's contorted face, close our ears to his cries of pain, watch his face carefully for a sign of an easing of the pain, look into his face, into his eyes, in sympathy, put an arm around his shoulder, and so on. (Cockburn 1994, p. 144)

According to Cockburn, Wittgenstein's remarks on the human form ought to be understood in connection with how he talks of our understanding of another person's pain as an *attitude* towards the other person and then also as a *response* to the other. Our responses of compassion often take certain bodily forms. We may look into the eyes of the other person or we may put an arm around his shoulder. In this sense Wittgenstein's remarks on the importance of the human bodily form differs from de Waal's remarks on the importance of our having physically similar bodies. While de Waal claims that physical similarity enables us to respond to other people in an analogical sense, Wittgenstein suggests that it is against a broad background of a shared life that the human bodily form is also of importance for the way we respond to

another's suffering, by for instance holding the person's hand or looking into his eyes. Cockburn (1994) further notes that we also often ascribe fear or other emotions to animals that physically can look and behave very differently from human beings. He maintains that our ascribing emotion to an animal is reflective of the way we can be concerned for the animal in question. That is, it is not because we observe certain physical similarities between human and animal behaviour that we draw the conclusion that the animal is afraid or happy. Cockburn suggests it is the other way around: "[...] our concern for creatures of a certain kind creates the possibility of our ascribing fear, pain and so on to them; and that, in turn, creates the possibility of observing a similarity between their behaviour and that of human beings (Cockburn 1994, p.150)." One can say the same thing about our ability to ascribe emotions to human beings. That we do so is in itself expressive of our attitude to the person, and expressive of how we are involved in each other's life.

1.6 Cooperation: emotivistic conceptions of human social life as instrumental

So far I have contrasted two theoretical perspectives on altruism; one rationalistic (transactional) and another that can be said to be emotivistic (and analogical) in kind. The rationalistic and emotivistic theoretical perspectives can appear to be quite opposite. However, in what follows I will argue that both perspectives suggest that human social life can essentially be described in instrumental terms, even

though they have somewhat different conceptions of this instrumental basis of human social life. I will argue that the tendency to consider human social life as having an instrumental essence is problematic.

Consider the following description by Jessica C. Flack and de Waal (2000) of food-sharing among chimpanzees:

Food sharing is known in chimpanzees [...] It is an alternative method to social dominance and direct competition by which adult members of a social group distribute resources among themselves. Most food sharing requires fine-tuned communication about intentions and desires in order to facilitate inter-individual food transfers. (Flack and de Waal 2000, p. 4)

Flack and de Waal further explain this food sharing:

[T]he reciprocity hypothesis—proposes that food sharing is part of a system of mutual obligations that can involve material exchange, the exchange of social favours such as grooming and agonistic support, or some combination of the two. (Flack and de Waal 2000, p. 5)

Flack and de Waal suggest that food-sharing is an example of what they call “the reciprocity hypothesis”, i.e. a system of exchanges of favours. Even though I do think Flack and de Waal’s emphasis on cooperative relationships has much truth in it, I think there are still some problems with the way they describe our care for each other as an exchange of favours. Such a description does not show how our practical, and non-practical, ways of living have meaning

as an *integral* part of our close relationships. Consider, again, what it means to eat together. For human beings, eating together is often a way of spending time together, of *being together*. We do not simply exchange food; we eat *together*. Eating together has a different meaning than eating alone. That we eat together often displays the self-evident way in which our lives are intimately woven together in multifarious ways. Sharing meals brings meaning to the whole situation. At the same time, the fact that we share meals is expressive of our sense of responsibility for the other. This is a responsibility that cannot be understood without acknowledging the importance of our shared presence for each other in eating together. This presence shows itself, for instance, in the fact that we often spontaneously talk with each other and listen to each other while we eat.

According to Flack and de Waal, communication is used in food sharing in order to “facilitate inter-individual food transfers”. I agree that talking while eating can be of practical importance. However, the fact that we talk at meals also often has a non-practical character. We simply talk because we are together. We are often spontaneously drawn into conversations while we eat with others. This may be something neither of us *decides* to do; and it need not have any further purpose. But it reflects how we can be there for each other. This spontaneous presence is in itself expressive of moral sensitivity towards the other; my ability to be drawn into the other person’s life, to become interested in his life and to take his life seriously.

Of course there can be situations where people do not talk with each other while eating. In some families it might be even more of a rule that they do not talk while eating. There can be various reasons for this. One reason can be that the married couple are deeply fed up with each other. In such situations a cold manner of talking merely in order to “facilitate inter-individual food transfers” might also take place. However, even if both animals and human beings sometimes share meals in a purely practical manner, and even if both animals and human beings sometimes communicate simply in order to exchange benefits while eating, such ways of eating are not any more basic or more natural than are our ways of enjoying eating together. Nor do such ways of eating and talking *explain* the multifarious ways we share a life with each other and find meaning in being together.

Wittgenstein writes: “Commanding, questioning, storytelling, chatting, are as much a part of our natural history as walking, eating, drinking, playing” (Wittgenstein 2001, §25). One thing Wittgenstein suggests here is that it is problematic to think of human nature as if it had an “essentially natural” part consisting of practical activities essential for survival, while our non-practical activities would not be as essentially natural but “cultural” or “psychological” features that enhance our practical cooperation. Another thing he suggests is that it is problematic to think of language as basically a practical tool. Wittgenstein mentions such things as chatting, storytelling and playing, that is, ways of talking and doing things that do not have a specific goal but are rather forms

of being with others. For Wittgenstein, our practical actions are not in any sense more basic or more natural than our non-practical ones. Nor is a practical use of language, such as exchanging information, more basic than a non-practical one such as chatting, or joking, or quarrelling or storytelling. The meaning of our practical activities can often not be separated from the way these are entwined in activities and ways of talking that do not have a practical character. A further thing Wittgenstein's remark suggests is that a large part of our activities have their meaning because these activities are integral to our standing in personal relationships. What it means for human beings to eat cannot be separated from the fact that we share meals and that it is common to talk spontaneously while we eat. In this sense our physical needs largely get their form and meaning from the fact that we live a shared life with others, and a shared life full of talk.

In *The Age of Empathy* (2009) de Waal seems, however, to acknowledge the fact that sharing a meal cannot basically be described as an exchange of favors (i.e. cooperation). He notes that apes not only help each other out when trying to get food but they also sit down and eat together. He asks:

[...] could it be that they just love to eat together? If both monkeys are rewarded, they will sit side by side munching on the same food. Do things taste better together than alone, the way we are more at ease having dinner with family and friends? (de Waal 2009, p. 113)

My answer here would be yes: sometimes food does taste better with family and friends and sometimes it tastes worse. De Waal then draws the following conclusion.

Perhaps it is time to abandon the idea that individuals faced with others in need decide whether to help, or not, by mentally tallying up costs and benefits. These calculations have likely been made for them by natural selection. Weighing the consequences of behavior over evolutionary time, it has endowed primates with empathy, which ensures that they help others under the right circumstances. The fact that empathy is most easily aroused by familiar partners guarantees that assistance flows chiefly toward those close to the actor. (de Waal 2009, p. 115)

Despite noting that apes (and humans) enjoy eating together, de Waal does not pursue the question what it implies; for instance, what it means to grow into a close personal relationship. De Waal assumes that close personal relationships are based on a cognitive and psychological phenomenon where our getting used to each other's company makes us more inclined to feel empathy for each other. This he sees as having a positive evolutionary effect since it enhances our capacity to cooperate.

Westermarck also emphasises the fact that people as well as animals enjoy being together. He claims that through evolution human beings and certain animals have slowly developed into beings that not only respond emotionally to each other and cooperate, but into beings that simply enjoy each other's company. He writes:

When gregariousness became an advantage to man, he would feel inclined to remain with those with whom he was living even after the family had fulfilled its object—the preservation of the helpless offspring. And he would be induced to do so not only from egoistic considerations, but by an instinct which, owing to its usefulness, would gradually develop, practically within the limits of kinship—the gregarious instinct.

By the gregarious instinct I understand an animal's proneness to live together with other members of its own species, apart from parental, conjugal, and filial attachment. It involves, or leads to, pleasure in the consciousness of their presence. The members of a herd are at ease in each other's company, suffer when they are separated, and rejoice when they are reunited. By actual living together the instinct is individualised, and it is strengthened by habit. The pleasure with which one individual looks upon another is further increased by the solidarity of interests. Not only have they enjoyments in common, but they have the same enemies to resist, the same dangers to encounter, the same difficulties to overcome. Hence acts which are beneficial to the agent are at the same time beneficial to his companions, and the distinction between ego and alter loses much of its importance.

But the members of the group do not merely take pleasure in each other's company. Associated animals very frequently display a feeling of affection for each other—defend each other, help each other in distress and danger, perform various services for each other. (Westermarck 1917, pp. 196-197)

The difference between Westermarck and de Waal is perhaps not a very big one but I still think there is a difference that is worth reflecting on. In the quote above Westermarck gives great importance to the fact that animals enjoy each other's company. He does not say that animals live together mainly in order to cooperate. Rather he maintains that the concept of cooperation gets meaning against the background of a shared life where we find it meaningful to live together for reasons that cannot be reduced to its being useful for us individually to cooperate. Thus cooperation cannot be understood as if it were a case of two strangers simply deciding to work together for the sake of some individual benefit to each. Nor can the concept of cooperation be understood as the basic aspect that gives meaning to all social relationships. Westermarck suggests it is the other way around. Our social relationships give meaning to what we mean by cooperation. My impression is that Westermarck touches on something important here, and that he also differs from de Waal. De Waal suggests that "natural selection" has enhanced our cognitive capacity for analogical imagination (i.e. empathy) which has enhanced our capacity to cooperate. In a similar sense Westermarck suggests that the origin of our "gregarious instincts" lies in the usefulness of cooperation. However, as I already noted, Westermarck also claims that our "gregarious instincts" give further meaning to how animals and humans cooperate. In this sense Westermarck leans partly towards de Waal's perspective and partly he differs from de Waal in the sense

that he does not reduce social relationships to a basically instrumental meaning.

In *The Ethical Demand* ([1956] 1997) K.E. Løgstrup distinguishes between mere cooperative relationships and the way helping each other is part of our standing in a close relationship, something he calls natural love.

People are bound to one another, among other things, through love, sympathy, and solidarity. By love and friendship they are bound together in an immediate way, whereas in solidarity they are bound together more through cooperative endeavor and common circumstances. But whether these ties are formed spontaneously or socially, it is these ties which constitute a person's existence. From them the individual acquires his or her life and its content. This is why every time one cares for another person in love, sympathy, or solidarity, he or she is him or herself rewarded through the maintenance of those relationships in which a person has his or her life and which constitute his or her existence. [...]

But the works of natural love are not on that account to be equated with works done for the sake of some favor given in return. On the contrary! It makes an indefinitely great difference whether that which benefits a person is seen as the maintenance of the relationships in which he or she has his or her life or whether it consists in the idea that 'one good turn deserves another.' If I help a person perhaps at every conceivable juncture, but always in a loveless manner and only because one good turn deserves another, this shows that that person does not belong to my life as an

indispensable part of it. The two of us are indeed of no concern to one another. Whether my favor to him or her is of help to him or her or not is in itself of no interest to me—so long as I receive his or her favor in return. He or she is only somebody who does me this favor, nothing more.

[...] it is not helpful but confusing to speak of reciprocity in connection with natural love, where both persons benefit from the deed by one for the other out of love. That I should benefit in return is of course not a condition or requirement that I attach to my love for the other person; it simply follows, so to speak, because indirectly the other person is a vital part of my life. (Løgstrup [1956] 1997 pp. 126-127)

Løgstrup's reflections are partly analogous with what Westermarck says about the "gregarious instincts". However, Løgstrup differs from Westermarck in the sense that he does not explain our caring for each other as dependent on "retributive emotions", that is, a cognitive disposition for analogical emotional responsiveness. Løgstrup argues that the practical help and care that we give each other is an integral part of a larger pattern whereby we acknowledge each other as persons. In this sense Løgstrup's reflections resemble Wittgenstein's. What it means to understand a child's physical needs cannot be separated from our acknowledging this child as someone to be with also in other ways, someone to eat together with, someone to chat with, quarrel with, and tell stories to, someone with whom we share a future life and a history. Løgstrup writes:

[...] the other person or persons by their mere presence function as a touchstone. Voice, tone, and gesture as such, however spontaneous, are always an unspoken invitation to the other person or persons to respond, to help, and to take them seriously. (Løgstrup [1956] 1997, p. 202, translation amended)

For Løgstrup, as for Wittgenstein, it is part of our natural life that we speak with each other and that we can stand in personal relationships. That we speak is, again, reflective of the fact that our relations in themselves contain an unquestionable responsibility for the other. Løgstrup suggests that this feature is also reflected in how we listen to the other's voice. He suggests that personal relationships, including what it means to help another and listen to another in such relationships, cannot be reduced to something basically instrumental.

1.7 Transaction: a rationalistic conception of human social life as instrumental

I shall now return once more to reflect on Cosmides' and Tooby's idea of altruism as transactions. I will maintain that both the emotivistic and the rationalistic perspectives on altruism are based on an instrumental conception of personal relationships. However, the conceptions of the instrumental form of human relationships differ.

Two features are central in Cosmides and Tooby's transactional perspective. First, human interpersonal encounters are described as if these were always based on calculation. Second, Cosmides and Tooby also describe our

transactional reasoning as taking place on a nonconscious level. This nonconscious level of reasoning they describe as a computational reasoning function that governs all our social engagement. They claim that all social engagement essentially takes the form of exchanges of benefits.

Cosmides and Tooby defend their view of a nonconscious transactional reasoning mechanism by referring to empirical evidence or examples from ordinary life. Some such examples were mentioned in the beginning of this chapter; “[...] how much chocolate you leave for j, how loud you play your music when j is trying to work, whether to clean up the mess or leave it for j [...]”. Surely we often think along these lines, but does this mean that there is an underlying nonconscious mental function of fitness reasoning taking place? It seems to me that these examples are described in a one-sided manner. One could also describe the examples in other, less calculating, ways. That is, by describing the examples in a certain manner Tooby and Cosmides create the impression that all our social engagement follows an underlying pattern of transactional reasoning. However, Cosmides and Tooby also refer to certain empirical observations of hunter-gatherer cultures, which they maintain will give us an idea of our unadulterated evolutionary past. With the example of how a “primitive” !Kung San woman reasons about food exchanges they want to point to the natural as well as universal forms of transactional reasoning in social situations.

When Agent X provides a benefit to Agent Y, triggering the expectation in both that Y will at some point provide a benefit to X in return, a social exchange relationship has been initiated. Indeed, within hunter-gatherer bands, many or most reciprocity interactions are implicit. (Cosmides and Tooby 2008, p. 72)

To illustrate their point, Cosmides and Tooby then offer the following quotation from Nisa, a !Kung San gatherer from Botswana who was interviewed by Marjorie Shostak¹⁶:

If a person doesn't give something to me, I won't give anything to that person. If I'm sitting eating, and someone like that comes by, I say, 'Uhn, uhn. I'm not going to give any of this to you. When you have food, the things you do with it make me unhappy. If you even once in a while gave me something nice, I would surely give some of this to you.' (Shostak 1981, p. 89)¹⁷

Commenting on this passage, Cosmides and Tooby then add:

Nisa's words express her expectations about social exchange, which form an implicit social contract: *If you are to get food in the future from me, then you must share food with me.* Whether we are San foragers or city dwellers, we all realize that the act of accepting a benefit from someone triggers an obligation to behave

¹⁶ The work by Shostak that Cosmides and Tooby refer to is *Nisa: The Life and Words of a !Kung Woman* (1981).

¹⁷ Shostak is quoted in Cosmides and Tooby (2008) p. 72.

in a way that somehow benefits the provider, now or in the future. (Cosmides and Tooby, 2008, p. 72)

I do not want to deny that these ways of reasoning about food sharing might be found anywhere among human beings, but the question is whether the example proves that there is an underlying nonconscious transactional pattern of reasoning in all social engagement.

Akhil Gupta and James Ferguson (1992) argue that Shostak's study on the !Kung people is problematic in several ways. One problem is that the image of a whole culture is based on a single person's words. They comment on Shostak's study: "The individual, Nisa, is granted a degree of singularity, but she is used principally as the token of a type: 'the !Kung.'" (Gupta and Ferguson 1992, p. 15) Gupta and Ferguson also question Shostak's suggestion that the !Kung people represent an untouched primordial people. They refer to an article by Mary Louise Pratt (1986) where she notes that the isolated life of the !Kung people is the result of centuries of colonialist violence. Gupta and Ferguson quote the following by Pratt:

What picture of the !Kung would one draw if instead of defining them as survivors of the stone age and a delicate and complex adaptation to the Kalahari desert, one looked at them as survivors of a capitalist expansion, and a delicate and complex adaptation to three centuries of violence and intimidation? (Pratt 1986, p. 49)

According to Cosmides and Tooby, anything that is done among hunter gatherers exemplifies the basic natural features of all human reasoning. Pratt's criticism of Shostak's study suggests, however, that the portrayal of "primitive" tribes can be misleading partly because it is taken for granted that such tribes have no history, that they have always lived like this. Gupta and Ferguson in turn argue that the image of the "primitive" tribe is problematic because its basis is too narrow (only one person in Shostak's case).

Not only is it problematic to take a single person's words to illustrate a whole culture's attitude. My impression is also that the !Kung woman's words can be understood in several ways. It is not self-evident that she is neutrally describing a general pattern of exchanges of benefits. To me it seems that the !Kung woman is embittered, as anyone can sometimes be. She seems to be fed up with sharing her food with others who sometimes apparently have taken advantage of her generosity. But the fact that people sometimes grow bitter and cynical does not prove that this is how we all nonconsciously "function" socially. The mere fact that she lives a life as a hunter-gatherer does not imply that her thinking is more natural and immediate than ours. Nor is the fact that all people around the world *sometimes* become bitter proof of the absolute basically transactional function of social life. There are also a lot of people in the world that are not bitter, and probably there are also !Kung San people who are not bitter all the time.

However, I might of course be wrong in my interpretation that the !Kung woman is embittered. Further, even if only one person is studied it might still be that she does mirror the general attitude among the !Kung people. Perhaps she is giving a neutral description of how she and others in her culture think about food exchanges. However, this does still not mean that the !Kung woman's words can be taken to show a general underlying natural pattern of human life. Even if researchers were to find a tribe that somehow was proven to have lived in a certain kind of completely "transactional" manner for thousands of years, it would still be problematic to maintain that it is therefore a pattern that explains the origin of all our social engagement. On the contrary, a people whose interpersonal encounters only took a transactional form would be a people very different from most other human beings in the world. It would be a people for whom much of how we approach each other would not have any meaning. So the problem is not only that it can be hard to find empirical evidence of an original "primitive" natural state of transactional life. I am suggesting that if such a form of life was found somewhere, it would no longer look like human life. By this I do not mean to say that transactions are not of central importance in human life in a multitude of ways. The problem arises when it is claimed that such ways of engaging with others are the essence of all social life.

Cosmides and Tooby's nonconscious transactional perspective is also connected with a predilection for economic jargon when describing human relationships.

They talk about “kin selection”, “welfare trade-off ratio” etc., when describing our care for family members. These expressions are expected to describe general underlying behavioural patterns. Our ordinary relational words such as “child” or “parent” are supposed to gain their meaning from these underlying strategic concepts. These expressions reflect, as I mentioned earlier, a conception of human relationships as based on calculation. They are supposedly relations that we choose to engage in (even if the choice may be on a nonconscious level). The idea is that we can choose to take any kind of attitude towards another person’s life, provided it enhances the chances of survival for the genetic information stored in us.

Løgstrup writes: “The other person is in such a real sense a part of our world that it is in fact awkward to refer to him or her as ‘the second person’ rather than as one’s child or spouse.” (Løgstrup [1956] 1997, p. 125, translation amended) According to Løgstrup, there is an ethical and relational meaning in the ordinary words we use when we talk about personal relationships. This meaning cannot be reduced to transactional relationships. The word “child” implies a human relationship where the child is involved in a certain way of living with others, sharing days and years in close relationships, and where there are people who talk with and care for this child. In this sense the meaning of the word “child” is integral to the self-evident form that our responsibility for the child takes. When Tooby and Cosmides use words such as “kin selection” they also reduce the meaning of personal relationships. From a perspective where all interpersonal relationships

are described as based on personal preferences it will not be possible to understand what it means to grow into a close relationship with another. Nor can our responsibility for one another be understood from such a perspective.

The words with which we, in everyday life, refer to “biological” relations and functions, such as eating, sex, fights, begetting children and caring for them, are imbued with meanings that go beyond both biology and strategic transaction. Consider a final example, what it means to say that a person is dying. This is not simply a neutral description of the person’s physical state. It is a moral description; when we say that a person is dying we also usually say that there is nothing more we can do to help him survive. At the same time, the fact that a person is dying and that we cannot help him to survive usually does not mean that we abandon him. The awareness of another person’s coming death often makes us attend to him in a special way. In particular it usually makes the dying person’s close ones attend to him. We try to make his last days as bearable as possible; we try to ease possible pain, we help practically, but we also talk, we share meals together, and we often share thoughts and memories about life and about loved ones. By this I do not mean that we necessarily always help our close ones or other people that are dying. Family relationships can sometimes be deeply damaged, filled with years of conflict that tear people apart. People can also wish for the death of another and yet kill others. However, these attitudes are not in any sense more basic for understanding human life than are the ways in which we often do care for others.

1.8 Conclusion

In conclusion, I shall reflect a little on the similarities and differences between the two perspectives on altruism that I have discussed. In one sense, there are clear differences between the rationalistic explanation of altruism that is given by Cosmides and Tooby, and on the other hand the emotivistic explanations that are given by Adam Smith, Darwin, Westermarck and de Waal. According to the emotivistic explanations human social life rests on a capacity for analogical imagination. According to the rationalistic explanation human social life rests on a capacity to reason in a calculating manner and make transactions. The emotivistic theorists are critical towards a conception of human beings as essentially calculating individuals. In this sense one can say that the emotivistic theorists, i.e. Smith, Darwin, Westermarck and de Waal distance themselves from the rationalistic perspective on altruism that is suggested by Cosmides and Tooby.

However, in another sense I have argued that the rationalistic and the emotivistic perspectives both rest on certain similar kinds of theoretical assumptions. I have claimed that the theories about empathy as a mental mechanism of analogical imagination do not actually contradict the view of altruism as a nonconscious, computational transactional mechanism. Rather one can think of these two explanations as existing on different levels. Cosmides and Tooby's perspective of human social life as based on a nonconscious algorithmic reasoning mechanism, does not necessarily contradict the theories of

empathy. To take an example, in one of the quotes earlier in this text de Waal states that we are not constantly “mentally tallying up costs and benefits [since these] calculations have likely been made for [us] by natural selection” (de Waal 2009, p. 115) This quite agrees with Cosmides and Tooby’s contention that the cost-benefit calculations appear nonconsciously.

Further, even though Adam Smith, Darwin, Westermarck and de Waal are critical of a transactional perspective in the sense that they are critical of the assumption that man is basically an egoistic being who only helps others in order to gain certain benefits; they do think that human social life originates from a need for cooperation. In this sense the rationalistic theorists as well as the emotivistic theorists maintain that human social relationships basically have an instrumental purpose. The emotivistic theorists are not actually opposed to a transactional perspective on the origins of human social life; they are merely opposed to *one kind* of transactional perspective. The theory about empathy can be seen as one form such a transactional explanation of human social life can take. It is an explanation that portrays transactions in a certain light, as involving cooperation between equals, and also as something that depends on a cognitive capacity for analogical imagination or analogical bodily responsiveness. However the emotivistic theorists also emphasise the social relationships that the individuals are born into, rather than a competition between antagonists. Cosmides and Tooby tend, on the other hand, to portray transactions as basically (nonconsciously) a competition between antagonists. I

have suggested that both these perspectives are problematic.

From my discussion above one might get the impression that I claim that human social life cannot be described from an evolutionary perspective. This has not been my intention to claim. I do agree with the above discussed evolutionary theorists that there are many social patterns of human life that enhance the survival of the species. However, that certain aspects of human social life are of evolutionary (and genetic) importance does not necessarily mean that these patterns can be explained by referring to a nonconscious algorithmic mechanism of social contract reasoning. By referring to a nonconscious algorithmic mechanism Cosmides and Tooby create a certain kind of reductive image of what they consider to be the natural form of human social interaction and also a certain reductive image of human understanding. They tend to describe human social interaction as if it always basically took a transactional form. I have argued that this impression is created by a one-sided use of examples of human interaction in combination with a one-dimensional use of expressions, in combination with a cognitive theory about human understanding.

I do also think it can, in certain contexts, be all right to say that human beings are cooperative animals. However, I have suggested that such a description can also have its limitations. A too general use of concepts like "cooperation" or "transaction" easily leaves out important aspects of what it means for human beings to share a life and thus also what it means to care for another. Further, I have

suggested that it is problematic to assume that our emotional responsiveness to each other has evolved as a useful cognitive tool that enhances cooperation. I have argued, by the help of Wittgenstein and Løgstrup that the meaning of the natural, biological, aspects of human life, including our practical ways of helping each other, in important ways is integral to how we stand in personal relationships.

Chapter 2: Theory of mind and infants' imitation of facial expressions

2.1 Introduction

As I mentioned in the introduction to this thesis, theory-of-mind theories can be divided into two more specific theoretical outlooks, one that is emotivistic in kind, and another that is rationalistic in kind. In this second chapter I shall focus on the conception of the human body that is mainly present in emotivistic theories. As I have mentioned earlier, the emotivistic outlook is also often called "*simulation theory*" while the rationalist outlook is called "*theory theory*". I do not generally use these expressions but in this chapter the distinction can be appropriate. According to simulation theorists, our capacity to understand other people rests on our ability to simulate other people's expressions and behaviour. Simulation theorists also maintain that this ability to simulate is largely innate. According to theory theorists on the other hand, our capacity to understand other people consists in a cognitive theorising function. Their suggestion is that this cognitive function enables us to actively construct a theory about other people's mental life on the basis of their behaviour. Theory theorists (or the rationalist strand) are then not as interested in the human body as are simulation theorists. However, despite a lesser interest in the human body among theory theorists, they are, I think, as entwined as simulation theorists in a problematic

conception of the human body, which is reflective of a problematic conception of interpersonal understanding. The following chapter is therefore not merely intended as a critique of simulation theory. Rather it is a general critique of the theoretical umbrella conception of theory of mind.

As I already noted, according to simulation theory, interpersonal understanding is largely considered to depend on a capacity for analogical imagination which again is considered to depend on the capacity to simulate bodily behaviour. These aspects have, among other things, been studied empirically in research on infants' capacity to imitate facial expressions. It is suggested that the infant's capacity to imitate other people's facial expressions is one of the primary routes for the child's development of interpersonal understanding. One of the central researchers in this field is the psychologist Andrew Meltzoff.

2.2 Empirical research on infants' imitation of facial expressions

Andrew Meltzoff's research on the infant's capacity to imitate facial expressions originates as a critical stance against the assumption that children are originally unaware of other people as well as unaware of the world. This has, in various forms, been suggested by among others Sigmund Freud, Jean Piaget and B. F. Skinner. Despite their great theoretical differences all these theorists assume that infants are originally unaware of other people. Meltzoff (2002) writes:

On classical views of human development, the newborn is cut off from others. Freud and his followers proposed a distinction between a physical and psychological birth. When the baby is born there is a physical birth but not yet a birth of the mind.¹⁸ The baby is like an unhatched chick within an eggshell, incapable of interacting as a social being because a 'barrier' leaves the newborn cut off from external reality. [...] Piaget's newborn is similar [...] Piaget¹⁹ claimed the baby is 'solipsistic.' The neonate has only a few reflexes to work with (e.g. sucking, grasping) and other people are registered only to the extent that they can be assimilated to these action schemes. The baby only knows his or her own actions. [...] Skinner²⁰ an animal behaviorist, gave his blank-slate infant even less to work with. One cannot really quote from Skinner about how children crack the puzzle of social cognition, because, in a sense, he does not think they ever do. Even adults are conceptualized as reacting to behaviors but not knowing the minds of their interactive partners. (Meltzoff 2002, pp. 7-8)

Though the three theorists mentioned in this passage are very different, they all suggest that children are originally unaware of the world and of other people. It is this general assumption of which Meltzoff is critical in his research.²¹

¹⁸ Meltzoff refers here to Freud (1911) and Mahler, Pine & Bergman (1975).

¹⁹ Meltzoff refers here to Piaget (1952c) and Piaget (1954).

²⁰ Meltzoff refers here to Skinner (1953).

²¹ I will not discuss the different theoretical positions between Freud, Piaget and Skinner in this thesis since their difference is not Meltzoff's

main focus. However, to illustrate the differences between especially Freud and Piaget I shall describe their theories briefly. Both Freud and Piaget suggested that children primarily live in a solipsistic state where they are unaware of the world as well as unaware of other people. Freud's theory on the development of understanding is empiricist in kind. According to him, the child is born into a chaotic world of sense impressions. The young child lives in a dream world that is formed through her own desires. Freud describes this state of infants living according to the pleasure principle as the infant being in an autistic state comparable with a bird's egg. Freud's conception of the child's sense impressions resembles the Cartesian idea that we essentially are caught inside our own sense impressions, where we cannot have certain knowledge about the outer world.

In contrast to Freud, Piaget ([1954] 1971) argued for the *logical* character of the child's development of interpersonal understanding and for the logical character of the child's development of understanding of the world. What it means to understand something, according to him, is dependent on our being able to see certain logical connections in the world, and what it means to understand another person shows in our being able to reason with other people in a logical manner. However, Piaget also considered it to be an important feature that the child's understanding develops through experience. Children's minds first put things together in ways that do not have to do with logic but with other factors, by association etc. Piaget calls this state "syncretic". Then slowly, through experience and practice, the child's understanding develops to a more logical way of functioning. David Hamlyn (1978) points out that Piaget takes a middle position between an empiricist perspective on understanding and a logical and rational perspective reminiscent of Immanuel Kant's conception. He considers an empiricist perspective as a primary level that precedes *true logical* understanding. The syncretic state of the child's mind means, according to Piaget, that the child is not truly aware of reality, or of other people, nor able to reason. Piaget writes: "During the earliest stages the child perceives things like a solipsist who is unaware of himself as subject and is familiar only with his own actions." (Piaget [1954] 1971, p. 397) Since interpersonal

Meltzoff has made a large amount of empirical research on the social responsiveness of infants. He claims that infants very soon after birth have a capacity to respond to other people's expressions by imitation. In this sense Meltzoff argues that infants are social from birth. However, even though I do think Meltzoff's critical stance against the assumption that children are solipsistic is important, I will argue that his research is based on a problematic conception of interpersonal understanding.

Since 1977 up till 2013 Meltzoff has conducted extensive empirical research on infant imitation. Meltzoff frequently refers to two of his own studies. In 1977 Meltzoff and M. Keith Moore studied infants' capacity to imitate facial expressions. In this experiment 12- to 21-day old infants were tested for whether they could imitate facial expressions such as tongue protrusion, lip protrusion and mouth opening. Meltzoff and Moore describe the test in the following way:

understanding, according to Piaget, consists in a capacity to reason in a logical manner, he concludes that infants originally live in a solipsistic state.

Piaget conducted a large amount of careful empirical research on infants. However, his conception of understanding as a matter of logical reasoning formed his ways of making empirical studies. It is striking that, even though his research is carefully made, he pays very little attention to how children are involved with other people socially from birth. This tendency to ignore infant's social responsiveness to others reflects his conviction that interpersonal understanding consists in a capacity to reason. Because of this, Piaget would probably not have considered Meltzoff's empirical research on infant imitation as proving that infants are not solipsistic.

Testing began with a 90-second period in which the experimenter presented an unreactive, 'passive face' (lips closed, neutral facial expression) to the infant. Each infant was then shown the following four gestures in a different random order: lip protrusion, mouth opening, tongue protrusion, and sequential finger movement (opening and closing the hand by serially moving the fingers). Each gesture was demonstrated four times in a 15-second stimulus-presentation period. This period was immediately followed by a 20-second response period for which the experimenter stopped performing the gesture and resumed a passive face. (Meltzoff and Moore 1977, p. 76)

The result of this test was that the infants generally did imitate the facial expressions. On the following page is a picture of Meltzoff and Moore's imitation test that was made in 1977.



Sample photographs from videotape recordings of 2- to 3-week-old infants imitating (a) tongue protrusion, (b) mouth opening, and (c) lip protrusion demonstrated by an adult experimenter. (Meltzoff and Moore 1977, p. 75)²²

Since the infants were capable of imitating several different facial expressions, Meltzoff and Moore suggest that these responses are not merely innate releasing mechanisms but “active matching processes”.

The hypothesis we favour is that this imitation is based on the neonate’s capacity to represent visually and proprioceptively perceived information in a form

²² The picture is taken from Meltzoff and Moore (1977). Reprinted with permission from AAAS: The American Association for the Advancement of Science.

common to both modalities. The infant could thus compare the sensory information from his own unseen motor behaviour to a 'supramodal' representation of the visually perceived gesture and construct the match required. In brief, we hypothesize that the imitative responses observed are not innately organized and 'released,' but are accomplished through an active matching process and mediated by an abstract representational system. (Meltzoff and Moore 1977, p. 78)

In 1983 Meltzoff and Moore²³ tested even younger infants, who were approximately 32 hours old (the youngest infant was only 42 minutes old). The result of this second experiment was that the infants tended to imitate the facial expressions such as tongue protrusion and mouth opening. Meltzoff and Moore claim that the results indicate that infants have an innate ability to imitate facial expressions.

Meltzoff and Moore's empirical research on infants' capacities to imitate facial expressions can be seen as an important criticism of especially Piaget's tendencies to neglect empirical studies on infants' responsiveness to other people. Meltzoff and Moore's research clearly illustrates that infants do respond to other people's expressions already at a very young age. In later studies Meltzoff and Moore have also tested whether infants can imitate after a delay of one day.

²³ See Meltzoff and Moore (1983).

Other research documents imitation after the memory delay is as long as one day. Six-week-old infants came in on one day, observed the gestures, and went home. They then returned the next day and were presented with the experimenter sitting motionless with a passive face. Infants successfully imitate based on their remembrance of things past.²⁴ If yesterday's adult had shown mouth opening, the infants initiated that gesture; if the adult had shown tongue protrusion, infants greeted him with that gesture. (Meltzoff 2002, p. 10)

Meltzoff also claims that the imitation research reveal that infants correct their efforts to imitate. "Research also reveals that the response is not rigidly fixed or stereotypic. Infants correct their imitative attempts so that they more and more closely converge on the model demonstrated." (Meltzoff 2002, p. 10) According to Meltzoff and Moore, these results indicate that the imitative responses are not merely spontaneous reflexes but can be considered as active efforts to match the other person's facial expressions. In this sense they suggest that the imitative responses display a cognitive capacity. A large amount of similar experiments have been conducted since the first studies on imitation, indicating that the results are robust. Meltzoff (2002) writes:

[...] the effect has now been replicated and extended in more than 25 different studies from 13 independent labs, including those from the US, England, Canada,

²⁴ Meltzoff refers here to Meltzoff and Moore (1994).

France, Switzerland, Sweden, Israel, Greece, Japan, and even in the rural Nepal.²⁵ (Meltzoff 2002, p. 11)

2.3 The influence from natural science on empirical research methods

From the empirical research described above it is clear that infants tend to imitate other people's facial expressions. However, even if the experimental results are interesting, there are some aspects that are ambiguous.

To begin with, there seems to be a discrepancy between on the one hand the empirical studies of facial imitation and, on the other hand, how children ordinarily encounter other people facially. An important aspect of the test is that it has a very simple form. The researchers make strong grimaces so that the infants are able to see these grimaces clearly and so that the infants are also able to respond to them without too much difficulty. Such procedures to simplify a test situation so that an infant is able to respond are in one sense all right. However, despite good intentions, it is not evident that the strong grimaces actually simplify the situation for the infants. One could ask whether this reflects an infant's normal way of being confronted with other people? Is it a normal situation that grown-ups show such strong grimaces to new-born infants or is it perhaps an unusual situation and thus perhaps a complicated and unnatural situation for the infant? Indeed parents may sometimes show strong grimaces to their child, but often it

²⁵ Meltzoff refers here to Meltzoff and Moore (1994) and (1997).

is also the case that parents do not show strong grimaces. However, even if we do not ordinarily constantly display such strong facial grimaces, this does not have to mean that the test is problematic. Something that indicates that the strong grimaces are all right is the fact that the infants do respond by imitating the facial expressions. Nevertheless there is an ambiguity regarding whether these tests actually simplify the situation or make it more complicated than an ordinary situation. One could also say that perhaps the children's capacity to respond *despite* the fact that the situation has little to do with how people would confront them in ordinary life suggests how very *sensitive* and alert children are to other people even in strange and complicated situations.

As I have suggested, there is a discrepancy between the empirical studies of facial imitation and how children ordinarily encounter other people facially, because ordinarily parents do not necessarily show strong facial grimaces to their new-born infants. How should this discrepancy be understood? One way to explain it could be to say that the infants' imitative reactions in the test situation points at a basic cognitive capacity that we can see most clearly, and most explicitly, in a limited experimental surrounding. The basic imitative capacity is, so to say, brought to the surface by reducing the normal life situation as much as possible and by giving "stronger" facial stimuli than normally. But is that actually what the tests reveal?

If we consider the character of the tests a little closer we can see certain further features that are problematic. One

such feature is that the studies of the infants' facial responsiveness tend to have a one-directional character, the researcher makes grimaces and the infant responds. The one-directional form of the tests appears to be reflective of a certain research practice deriving from natural science. According to Urie Bronfenbrenner (1978), there is a deep-rooted tendency in psychological research to construct test situations in such a way that only the single individual's responses are studied. What is left out then is often the fact that an individual's responses are entwined in a pattern where two persons *both* respond to each other. In *The Ecology of Human Development* (1978) Bronfenbrenner writes:

The thesis that behavior in dyads is generally reciprocal is widely accepted in theory, but it is often disregarded in research practice. The failure to take two-way processes into account reflects the inertia of the traditional laboratory model with its classical participants—an experimenter, identified cryptically as E, and another person equally informatively described as S, the subject. The term subject is apt, for with few exceptions the process operating between E and S is viewed as unidirectional; the experimenter presents the stimulus, and the subject gives the response. Of course in theory the influence can occur in both directions, but once the researcher puts on the white coat of scientific invisibility, she tends to focus solely on the behavior of the experimental subject [...] (Bronfenbrenner 1978, p. 61)

Bronfenbrenner's suggestion is that a certain kind of traditional research practice tends to lead researchers to focus on the individual even if in theory they might acknowledge that human behaviour must often be seen as *interactive* responses between *two or more* individuals. Bronfenbrenner does not discuss the above-described tests on imitation, but I think his thoughts also point at a problematic one-directional form in these tests. Even if the original research interest is to find out how children learn to *communicate*, this research on "communication" leans towards a one-sided focus on only the *child's* responses. It is unclear then in what sense such a focus can reveal patterns in the development of communication.

However, Meltzoff and Moore's tendency to construct a one-directional experimental situation has not merely to do with their being influenced by certain research methods that one finds in natural science. There is also another reason why they have built up the experimental situation in a one-directional manner, which I shall discuss below.

2.4 The influence from theory-of-mind theory on empirical research methods

Meltzoff (2002) suggests that the infant's capacity to imitate is connected with the development of a theory of mind. He writes:

Imitative experience with other people serves as a 'discovery mechanism' for social cognition, engendering interpersonal understanding that outstrips the innate givens and leads to empathy,

perspective-taking, and theory of mind. (Meltzoff 2002, p. 7)²⁶

Meltzoff continues:

[...] imitation is based on [the] infant's capacity to register equivalences between the body transformations they see performed by other people and the body transformations they only feel themselves make. On this account, facial imitation involves crossmodal matching. Infants can, at some primitive level, recognize an equivalence between the acts they see others do and the acts they do themselves. There appears to be a very primitive and foundational 'body scheme' that allows the infant to unify the seen acts of others and their own felt acts into one common framework. (Meltzoff 2002, pp. 12-13)

A bit later Meltzoff continues:

[...] intentions underlie and cause bodily movements, and reciprocally, one can read intentions from body movements. But the intentions themselves are not directly seen, heard, tasted, or smelled. The developmental problem is clear and irresistible: Is there any evidence that infants read below the surface

²⁶ It appears that Meltzoff has gradually adopted the idea that infant imitation is a theory of mind function. In Meltzoff and Moore's articles from 1977 and 1983 this is not yet explicitly suggested. The gradual theoretical development towards an explicit theory-of-mind theory, including simulation theory, in Meltzoff's research might be connected with the fact that theory-of-mind theories have gradually gained in popularity.

behavior and understand the intentions that lie behind them? How do they come to this interpretation of bodily acts? (Meltzoff 2002, p. 17)

In another article Meltzoff (2010) writes:

Although we directly observe other people's behavior, we interpret people to be more than their biological movements. We think of them as having internal mental states just like us. We think that human beings want, think, and feel, and that these states lead to their actions. Our ideas about these mental states play a crucial role in our interactions with others and the regulation of our own behavior. (Meltzoff 2010, p. 15)

Meltzoff (2013) describes the essence of his theoretical question in the following way: "We know ourselves from the inside and others from the outside: How do we understand what it is like to be another person, to feel what the other person feels?" (Meltzoff 2013, p. 140) According to Meltzoff, interpersonal understanding consists in the capacity to "interpret" people's bodily behaviour, i.e. to see that people are "more than their biological movements", to see that people have "mental states". Briefly one can describe this conception of interpersonal understanding as epistemological in kind, and also body-mind dualistic in kind. In order for a child to learn to understand others, the child must learn to see that other people are *intentional* beings. It is, according to Meltzoff, through imitation that infants learn to interpret other people's bodily behaviour, so that they eventually

gain knowledge about other people's intentions. Meltzoff further suggests that by imitating others, the infant comes to recognize that other people are "like me". This experience of similarity, according to him, is a central aspect of the infant's primary social encounters. He writes:

The child, even the newborn, processes the movements of other people and recognizes: 'that looks the way this feels' or 'those acts are like these acts.' the fact that others are seen as 'like me' provides an interpretive lens for infants' first social encounters. (Meltzoff 2011, p. 52)

My suggestion so far is that one can discern two reasons for the one-directional character of Meltzoff and Moore's empirical tests on infants. Partly the one-directional character of the tests reflect what Bronfenbrenner describes as the "traditional laboratory model" where the researcher observes the subject's responses. However, partly the one-directional form of the test also grows out of the dualistic philosophical supposition that a child originally perceives other people as physical bodies. Further, the strict focus on imitation in the test arises from the assumption that the only way a child learns to understand others is through the use of an analogical method of behaviour which eventually enables the child to decipher other people's intentions. In this sense there seems to be a double influence from on the one hand a certain kind of classical (natural scientific) research practice and on the other hand a philosophical theory. It is hard to say which of these influences is stronger, but both influences support each other.

As I already suggested, one central feature with the imitation experiments is the idea that facial expressions have a very rigid, stable and also a very marked character. The expressions are rigid and motionless; a tongue sticks out, or a mouth is wide open. In his paper "What's in a Smile?" (2009) Lars Hertzberg claims that our facial expressions actually do *not* have a rigid form. Facial expressions have a changing and fleeting character. And they are usually not very marked.

[...] smiles [...] do not constitute disparate units, smiles form part of a continuum of human expressions. Part of what is involved in smiles being [...] a form of bodily openness, is that they are a living, organic feature of the human face. (Hertzberg 2009, pp. 120)

Hertzberg also quotes Wittgenstein who talks of it being difficult to understand such a rigid motionless facial expression as a smile:

[...] a facial expression not susceptible of gradual and subtle alterations; but which had, say, just five positions; when it changed it would snap straight from one to another. Would this fixed smile really be a smile? And why not? (Wittgenstein 1980, §614)

Commenting on this passage, Hertzberg then adds:

A face which switched from neutral to a broad grin and then back again without transitions would not really come across as smiling or indeed as expressive in any way, rather it would strike us as undergoing some

strange contortions. The face would not *be alive* with the smile. (Hertzberg 2009, pp. 120-121)

If facial expressions do not have this kind of rigid form in real life, then it is questionable whether the imitation experiments reveal how a child grows into responding to others and understanding others.

The conception of facial expressions as consisting of rigid movements is connected with a certain way of talking about the human body in the test, as if the body was a mere surface. Meltzoff talks about children imitating “surface behaviour”. This use of the word “surface behaviour” could of course merely reflect a kind of professionalised way of speaking in a scientific research context. However, it is not merely a manner of speaking but reflective of the dualistic assumptions that form the tests. The problem here lies in the assumption that people essentially consist of a physical “outer” body and an inner mind. Thus we basically see bodies per se as outer objects. The use of the words “surface behaviour” or “imitation” are also reflective of the before-mentioned tendency merely to focus on the single individual’s reaction. The researchers have built up the test so that only the children are studied. What is *not* studied is the *interaction* between child and parent.

A somewhat similar approach as in the above described research on children can be seen in Paul Ekman’s research on facial expression. Ekman’s studies on facial expressions began in the 1970’s. Since then he has made extensive empirical studies. I will not discuss Ekman here at length

but I want to point out that he has a similar one-directional approach in his research as Meltzoff has in his research. This is connected with the fact that he has a similar conception of interpersonal understanding.

Ekman and Wallace V. Friesen (1971) studied so-called “primitive” tribes in Papua New Guinea. Their aim was to study whether facial expressions have a similar meaning across cultures. Ekman and Friesen conducted their study by showing photographs of facial expressions to the people in the tribes. The result of the study was that the people in the Papua New Guinean tribes generally attributed similar emotions to the pictures as Western people had attributed to the pictures. Since this empirical research in 1971 Ekman has conducted an extensive amount of further empirical research on facial expressions, which has made him very influential.

However, even if the focus of Ekman and Friesen’s study differs from Meltzoff and Moore’s studies, one can say that the studies are built up in a similar one-directional manner. As I already noted, Ekman and Friesen showed the people in Papua New Guinea pictures of facial expressions that the people were then requested to *define*. They did not consider how facial expressions can have meaning for people who are involved in some sort of *interaction*. Ekman and Friesen’s study of facial expressions also rests on epistemological and body-mind dualistic conceptions of interpersonal understanding. In an article from 1993 Ekman writes:

It is the morphology, the momentary configuration produced by the contraction of a particular set of facial muscles, that provide the information about whether it is anger, fear, disgust, sadness surprise, or enjoyment. The dynamics of the movement also contains additional information about the strength of the emotion and whether it is genuine, although that information is also signalled morphologically. (Ekman 1993, p. 389)

Ekman describes bodily expressions as “providing information” in the form of physical signs.²⁷ His description suggests a dualistic conception of human beings, it suggests that facial expressions reveal something invisible inner. This epistemological and body-mind dualistic conception of interpersonal understanding forms Ekman and Friesen’s empirical research in a similar one-directional way as Meltzoff and Moore’s research on infant imitation.

In a similar sense as Meltzoff and Moore who study infants’ responses to faces, Ekman assumes that facial expressions are something we primarily *observe* rather than something we respond to in our daily life. And he also assumes that bodily expressions have a rigid on/off form that we can label. It is also assumed that it is *easier* to see the human body as a physical surface than it is to see

²⁷ Ekman was inspired by Charles Darwin’s research on facial expressions. To a large degree Ekman’s studies also resemble Darwin’s studies. See Darwin (1872), *The Expression of Emotions in Man and Animals*.

meaningful expressions. Frank Ebersole (2001) describes this tendency to consider the body as a kind of physical shell:

[...] bodily movements are to be simple, easy to see. They are common and can be pointed out, anytime and anywhere. They are conspicuous items in any inventory of the physical world. (Ebersole 2001, p. 371)

The whole idea of facial expressions as outer signs that we *observe* in others, distorts the role of expressions in human life. Hertzberg (2009) writes:

The concept of a smile and what it may signify does not enter our life through a discovery. Neither have smiles been introduced as a social convention. The word 'smile' is learnt in a setting in which smiles already have a place, in which people smile and respond to smiles without giving the matter much thought. Both the view of smiles as natural phenomena and the view of them as applied conventions, then, are expressive of the same misconception: the idea that smiles can be studied as physical configurations without regard to their significance. (Hertzberg 2009, p. 123)

Cockburn (2009) writes:

Bodily expression, and in particular facial expression, has a central place in our interactions with each other: is a central form of our *contact* with others. I respond to the other's smile of pleasure or amusement with a smile. I turn away from the other's angry gaze, or return her friendly or loving gaze. I shrink in the face of

the other's manifest hostility or anger, and the other may respond to my shrinking, perhaps with a softening of her expression; as, more generally, she may respond to my timidity with an encouraging smile. In many cases this contact involves an acknowledgement of a world that we share with the other. We exchange a smile at the remark just made, a look of surprise when the guests arrive on time, or a fearful glance when we recognize the danger. (Cockburn 2009, p. 129)

Both Hertzberg and Cockburn suggest that it is problematic to consider the meaning of facial expressions as if they basically consisted of certain physical movements that we observe and learn to interpret. They argue that facial expressions have their meaning as part of our mutual responsiveness and then also as part of what it can mean to stand in a personal relationship.

My aim so far has *not* been to say that imitation has no importance at all for how infants learn to understand others. Neither have I wanted to suggest that it has *no* importance at all what kinds of physical movements may occur in a person's face when he or she smiles or is sad or angry etc. Clearly infants do sometimes imitate facial expressions, and clearly one can also discern certain general physical patterns of facial expressions when people look happy, sad, angry etc. What I have wanted to question is the idea that the infant's capacity to imitate facial expressions would be the *basic cognitive capacity* that enables the infant to interpret *other minds*. That is, I have wanted to question a tendency to consider imitation as the sole (or most important) "method" that enables us to

understand others. I have also questioned the idea that the meaning of our facial expressions can, in some essential sense, be reduced to the mere physical movements of the face. These ideas are based on epistemological and body-mind dualistic conceptions of interpersonal understanding. They are based on the idea that other people have outer bodies with inner invisible minds about which we need to gain knowledge.

2.5 What do we mean by imitation

I shall now return to reflect a bit further on what Meltzoff means by imitation. According to Meltzoff, the infant's tendencies to imitate other people's facial expressions will eventually enable the infant to figure out the intentions of other people. As I have already argued, Meltzoff and Moore's empirical research is more unclear than it can seem at a first instance. I have claimed that the empirical research has a one-directional character. Partly this is reflective of the influence from natural scientific research practice. Partly it reflects an epistemological and body-mind dualistic conception of interpersonal understanding. There is, however, a further unclarity concerning what Meltzoff means by "imitation". The unclarity is connected with his supposition that there are some such general inner mental things as "intentions", and that there is one method, i.e. imitation of "bodily movements", for understanding these "intentions".

I noted earlier that according to Meltzoff the results of the experiments indicate that infants' imitation of facial

expressions is not merely a reflex but an active effort. This further indicates, according to Meltzoff, that infants have a cognitive capacity for mindreading. Meltzoff, Rebecca A. Williamson and Peter J. Marshall (2013) further connect the infant's capacity to imitate facial expressions with the capacity to learn to imitate other people's actions and thereby to learn to do things.

Before language becomes available to the child, imitation is a chief mechanism by which they learn about tool use and acquire causal knowledge about how novel objects and machines work. This 'instrumental imitation' continues to play such a role in adults: how to tie a knot, build a fire, or use a lever is more efficiently learned through studying other's behavior than via an instruction manual or a linguistic narrative. (Meltzoff, Williamson and Marshall 2013, p. 287)

Indeed learning to do things often involves a great deal of imitation. But such imitation often *presupposes* that we are *already* acquainted with a way of living where certain ways of doing things are seen as meaningful intentional actions, not merely as physical movements. Think for instance of how to learn to skateboard. In order to be able to imitate specific skateboarding moves you need to understand what is being *done* and *who* is doing something and who is *not* doing something. Without knowing the difference between a person walking his dog and a person skateboarding you will not be able to see whom you should imitate. If you merely see moving physical bodies

any bodily movement becomes relevant or irrelevant. Should you hold a leash and pull the furry animal? Or should you do something with that flat thing on wheels? Perhaps the skateboarding girl takes a little pause to catch her breath, drinks a bit of water from a bottle, scratches her head, picks her nose. Are *these* the bodily movements you should imitate? The point here is that usually when we talk about one person imitating another person's bodily movements this already presupposes that the person who imitates sees a meaning in what is being done. We do not see the skateboarding girl's bodily movements as *first* "purely bodily" and *then* "intentional", but from the first jump we see them as meaningful and intentional, and that is why we can focus on what she is doing and also try to learn to do the same thing. Our ability to focus on another person's bodily movements entails that we conceive her or his actions as having a meaning.

What one can see from this example is that it is important to distinguish between the sort of imitative action that requires that we already see a meaning in what another person does, and the kind of spontaneous imitative response that an infant will sometimes display. The spontaneous imitative response that infants exhibit have little to do with the kind of imitation I have described above. The kind of imitation described above is an advanced capacity for action requiring a conception of various human contexts, and requiring a capacity to control and focus on one's own bodily movements. It is not a matter of merely imitating physical movements in general. This does not mean that the child's spontaneous

responsiveness is unimportant or that we could not say that it is a kind of imitative response, but it means that the child's response cannot be described as some *general* cognitive capacity that enables the child to interpret other people's "intentions" by imitating their "bodily movements". Even if I do agree with Meltzoff that learning has much to do with imitation, I have argued that his conception of learning is still based on a problematically general and inner conception of intentions, which is connected with a problematic conception of imitation as well as with a problematically physical conception of bodily movements. He suggests that there is some such general thing as bodily movements that we first perceive as mere physical movements, which then enables us to interpret the person's intentions. I have argued that our capacity to see a meaning in what people do, that is, our capacity to see that someone is, for instance, skateboarding, does not rest on a primary capacity to observe other people's mere physical movements, which we then imitate and then interpret as having a certain meaning. Learning to *observe* other people's bodily movements is something we do when we are older, it is not a basic form our learning to understand each other takes. And learning to observe other people's *physical* movements is an even more specialised capacity; something that you may learn to do as a doctor.

2.6 A dialogical perspective

Another way to think of a child's development of interpersonal understanding can be found among proponents of so called *attachment theory*. A central idea in attachment theory is that a child's development of interpersonal understanding grows out of a close *reciprocal* relationship with the caretaker. In his book *Attachment and Loss* ([1969] 1997), the psychologist John Bowlby writes:

The patterns of interaction that gradually develops between an infant and his mother can be understood only as a resultant of the contributions of each, and especially of the way in which each in turn influences the behaviour of the other. (Bowlby [1969] 1997, p. 204)

A bit later Bowlby continues:

Very frequently, it is found, babies behave in such a way that they maximize the kinds of stimuli that emanate from humans. Examples already given include a tendency to look at pattern, or at least contour, especially when it resembles a human face, and a tendency to listen to a human voice, especially a female one, and to cry when it ceases. Another bias, present from very early days, is a tendency to look at anything that moves in preference to something static.

Not only are babies biased to behave in special ways towards humans but mothers are also biased to behave in special ways towards babies. By bringing her baby into a face-to-face orientation to herself a mother gives him opportunity to look at her. By cradling him to herself in a ventro-ventral position she is likely to

elicit reflex responses that not only orient him more precisely to her but also give him the chance to use mouth, hands, and feet to grasp parts of her. And the more each experiences the other in these interactions the stronger do the relevant responses of each tend to become. In this reciprocal way early interaction between mother and baby is begun. [...] Whilst being breast-fed, a newborn who happens to be alert and to have his eyes open will often fixate his mother's face [...]. (Bowlby [1969] 1997, pp. 271-272)

Bowlby's perspective differs from the perspective put forward in imitation research. He does not deny that children can have certain natural tendencies to imitate, but he does not maintain that this is the fundamental aspect leading to interpersonal understanding. On the contrary, Bowlby emphasises the *mutual* interaction between parent and child. It is not only the child that looks at the parent's face and reacts, but also the parent actively engages the child and enables the child to look at him or her.²⁸ Bowlby's

²⁸ It is important to note that attachment theory has been criticized. In her book *Mass Hysteria: Medicine, Culture and Mother's Bodies* (2005) Rebecca Kukla discusses tendencies in attachment theory to glorify and essentialise the relation between mother and child. One form this takes is through making it look as if the mother-child relation, and especially breastfeeding, is the essential basis for all social relations. Kukla questions tendencies to talk about breastfeeding as if it were the absolutely most essential and most natural form of interpersonal and bodily contact with a child, and a contact that is fundamental to all other forms of interpersonal contact. According to Kukla, the emphasis on the mother-child relation and breastfeeding that can be seen in attachment theory has old roots deriving from Rousseau and the period of the

perspective on what it means to simplify a situation differs in an important way from the conception of how to simplify a situation that is discernable in the imitation test. While the imitation test is built on the presupposition that

European Enlightenment. Rousseau argued that the natural order of social life derives from the natural relation between child and mother. However, Rousseau does not simply try to explain the origins of social life, but his reflections have political aims. He connects the mother-child relationship with a natural sense for "the fatherland". In this sense his thoughts are not merely meant as a social theory of interpersonal development but also play a political role during a time when, among other things, the concept of the nuclear family was gaining importance. According to Kukla, this trend can also be seen in attachment theory. I find Kukla's critical reflections to be important. It is important to be aware of the risk of constructing pictures that through the invocation to human nature may be used politically. However, this does not mean that the dialogical perspective in attachment theory is essentially flawed. Nor does it mean that we should instead stick to theory of mind. It is also important to see that the supporters of attachment theory comprises a broad range of theorists of whom some are clearly propounding a certain political agenda according to which mothers should stay at home with their children. Bowlby is not, as far as I can see, doing so. Bowlby notes the following about his use of the word "mother". "Although throughout this book the text refers usually to 'mother' [...] it is to be understood that in every case reference is to the person who mothers a child and to whom he becomes attached. For most children, of course, that person is also his natural mother." (Bowlby [1969] 1997, p. 29) Bowlby's thinking derives from a time when mothers were the primary caretakers of infants. Therefore he generally talks of mother instead of for instance "caretaker" or some other more neutral concept. But he does not argue that mothers are the only ones who can have a genuine relation with a child. Nor does he maintain that mothers should stay at home.

strong grimaces will make it easier for the child to respond, Bowlby emphasises that the parent actively puts the child in such a position that both can look at each other during such pleasant situations as for instance feeding. In this sense there is a form of *care* involved in how the parent tries to make things easier for the child, a care that does not consist in showing exaggerated meaningless grimaces, but a care that shows in how the parent enables the child to sense a certain kind of *meaning* in the presence of the parent. This can take both positive and negative forms. Positively it can be a way of enabling the child to relax, where the relaxing can take the form that both look into each other's eyes. Negatively a tired and fed up parent can, for instance, handle a child harshly when changing diapers in the middle of the night.

Instead of studying a child's responses towards facial expressions made by a stranger for a brief moment, Bowlby studies the child's growth into a close and meaningful long-term relationship and interaction. These are relationships that largely take the form of bodily closeness, and that consist in an everyday way of being together, and these are relationships that develop and deepen through patterns of interaction (patterns that change and develop) that span over many years.

The presence to each other is also something that concerns the voice. Bowlby writes:

Not only is mother an interesting and rewarding object to watch but she is an interesting and rewarding object to listen to. [...] it seems probable, too, that, just as in

the case of visual attention and tracking, a baby's auditory attention and pursuit are encouraged and augmented by processes of feedback and learning. On the one hand, her infant's interest in her voice is likely to lead a mother to talk to him more; on the other, the very fact that his attention to her has the effect of increasing mother's vocalisations and other baby-oriented behaviour is likely to lead the baby to pay even more attention to the sounds she makes. (Bowlby [1969] 1997, p. 274)

Again Bowlby emphasizes the interactive pattern of how the voice comes to have meaning for the child. It is not merely that the parent tries to make the child respond correctly to the voice. Rather it is as if the attention that *both* pay to each other's presence also enhances the spontaneous responsiveness of *both*, it enhances the parent's tendency to talk to the child as well as the child's tendency to listen to the parent's voice. Parents talk to their child long before it can understand any words. This is not generally a matter of the parent trying to make the child imitate words so that it eventually recognizes the meaning of these words. Rather the talking is often simply a way of being together, it is a form of presence in the various practical things that are done.

Bowlby brings forth dimensions of how a child comes to respond to facial expressions (or to the voice) that cannot be discerned from the perspective represented in imitation research. In imitation research the adult does not adapt his behaviour to the child's responses. It is just the child that should learn to imitate the adult. And even if the

experiment in some sense is simplified, the simplification has very little to do with trying to communicate with the child and trying to make a situation more meaningful and comprehensible to the child. The researchers are merely trying to produce certain responses to facial images or to facial movements. They have, so to say, from the start decided that infants are not possible to communicate with, because it is taken for granted that infants can only see physical surface images and imitate them. Clearly parents often address children in a kind of simplified manner, but this is done so that the child can by and by come to sense a meaning in the situation.

That we respond to another person is so deeply part of what it means to acknowledge another that it can feel very awkward not to do so, even to an infant. In *How Infants Know Minds* (2008) the psychologist Vasudevi Reddy describes a test where adults were requested to display a completely still and unresponsive face to an infant. Researchers then studied the reactions of the infant. The result was that after a little while the child became unhappy. But also the adults found this situation of unresponsiveness as emotionally difficult.

Confirmation or recognition of the other can happen—or not happen—in many ordinary ways. In all the perturbation experiments, the adult (whether the parent under instruction to hold a still face or the experimenter manipulating the video play) is explicitly not confirming the infant—not acknowledging or recognising the infant's previous acts or the infant herself. Mothers and other adults asked to engage in

still-face experiments sometimes report finding them emotionally difficult. And this is why: they are being asked to act as if the infant isn't there—to not acknowledge the infant. (Reddy 2008, p. 84)

Something similar is going on in the imitation tests. The adult is not actually responding to the child but merely trying to produce a certain response from the child. D. W. Hamlyn writes in *Perception, Learning and the Self* ([1983] 1994).

[..] early learning is very much a function of personal relationships that exist between the child and other human beings; without this or something like it, it is difficult to see how learning could go on at all, let alone make much progress. (Hamlyn [1983] 1994, p. 144)

How should one then explain the occasional imitative responses that infants can exhibit? Instead of considering a child's imitative responses as displaying a primary mindreading function, the child's tendency to imitate facial expressions can be seen as part of a spontaneous dialogical engagement with others. Children are intensely awake to other people. Certainly this might well to a large degree be something innate but it is also something that deepens and changes as the child's interaction with others deepen. It is a responsiveness that is entwined in how parents are responsive to the child. Imitation is merely *one* occasional pattern that their dialogical presence to each other can take. It is one aspect of how the child can be engaged with another. These ways of responding do not have to do with

learning to read minds, though they are one aspect of how we grow into a close relationship with others and into an everyday way of being with others.

Admittedly, Meltzoff (2013) does acknowledge the fact that imitation can be part of a dialogical relationship. He notes that imitation can often be a way of playing with a child.

[...] mutual imitation games deepen a sense of relationship. Mutual imitation indicates ‘communing’ or ‘being with’ someone else, even prior to the time that linguistic exchanges are possible. [...] caretaker’s mirroring serves the functions as a physical mirror. Infants can use imitative interactions to learn what the self looks like. [...] Through such social mirroring, infants gain a better sense of what their own felt acts look like. (Meltzoff 2013, p. 141)

Here Meltzoff acknowledges the dialogical and relational character of imitation games. In this sense Meltzoff here appears not to have a purely epistemological or a purely dualistic conception of interpersonal understanding. That is, he does not describe the purpose of imitation as if it merely had the function of enabling the child to gain knowledge about the other person’s inner intentions. However, Meltzoff still seems to consider the “deepened sense of relationship” as a secondary feature. He suggests that the mirroring enables the child to gain an understanding of what its own acts look like. In this sense Meltzoff still considers interpersonal understanding,

including self-awareness, as dependent on a cognitive analogical mechanism.

There is also a further problem with Meltzoff's emphasis on imitation. Even though he notes that imitation games can deepen the relationship between parent and child, he does not note that there are a lot of *other* things that children and parents do together, where imitation is not the central feature. Meltzoff's emphasis on imitation as the basic source of interpersonal understanding suggests that he regards interpersonal understanding as consisting of one basic mental function of analogical imagination. However, even though I agree that infants sometimes imitate others I think it is important to acknowledge that there are a lot of other things that parents and infants also do together. Meltzoff appears to suggest that imitation could be considered meaningful in itself, regardless of the larger life-context that the child is part of, and regardless of other kinds of responses displayed by the child or by the parent. Göran Torrkulla (2009) writes: "[...] the child, from its very birth, is *met* with the *full* range of expressive possibilities of the community into which it is born." (Torrkulla 2009, p. 144) It is only against a background of daily mutual engagement in all sorts of *different* ways that the idea of an infant occasionally imitating another is comprehensible. A child's imitative response to another person gets its meaning from the way the responses are entwined in a broad web of daily responses that do *not* take the form of imitation. The point is that surely a child might occasionally imitate other people, but this response cannot be understood if we isolate it from the rest of her

life, or if we isolate this feature from how adults respond to the child in many *different* ways.

2.7 Mirror neurons

As I have suggested, it is not self-evident that Meltzoff's perspective contradicts Bowlby's perspective. Both can be taken to support a perspective on children where the child's growth of interpersonal understanding is dependent on its entering into social relations with others. Meltzoff does also, in his later works, occasionally note that imitation is part of how parent and child can play together. In this sense his theory of imitation can also seem to resemble Bowlby's dialogical perspective. However, Meltzoff mainly discusses this feature on another, allegedly more fundamental and more technical level than Bowlby does.

Research has also been done on so-called "mirror neurons", and this research could also be taken to be compatible with, though moving on a "deeper" level than, Bowlby's more socially oriented dialogical perspective. The theory about mirror neurons consists in the idea that our capacity to respond emotionally and physically to other people's actions and emotions is based on a certain kind of neural system that "mirrors" other people's emotional or behavioural reactions. The "mirror neuron" phenomenon has also been studied empirically in monkeys. In their paper "Grasping the Intentions of Others with One's Own Mirror Neuron System" (2005) the neuroscientists Marco Iacoboni, Istvan Molnar-Szakacs, Vittorio Gallese,

Giovanni Buccino, John C. Mazziotta and Giacomo Rizzolatti write:

Mirror neurons are premotor neurons that fire when the monkey performs object-directed actions such as grasping, tearing, manipulating, holding, but also when the animal observes somebody else, either a conspecific or a human experimenter, performing the same class of actions. (Iacoboni, Molnar-Szakacs, Gallese, Buccino, Mazziotta and Rizzolatti 2005, p. 0529)

In another article Rizzolatti reflects on the possible importance of mirror neurons.

What is the functional role of the mirror neurons? Various hypotheses have been advanced: action understanding, imitation, intention understanding, and empathy.²⁹ In addition, it has been suggested that mirror-neuron system is the basic neural mechanism from which language developed.³⁰ (Rizzolatti 2005, p. 419)

Rizzolatti suggests that the “mirror neuron system” in our brains is the root of all our social engagement, including our capacity to talk. In their article “Language within our Grasp” (1998) Rizzolatti and Michael A. Arbib write: “It is likely that the human capacity to communicate beyond that of other primates depended on the progressive

²⁹ Rizzolatti refers here to Rizzolatti and Craighero (2004) and to Gallese, Keysers and Rizzolatti (2004).

³⁰ Rizzolatti refers here to Rizzolatti and Arbib (1998).

evolution of the mirror system in its globality.” (Rizzolatti and Arbib 1998, p. 192)³¹ Meltzoff (2010) argues that the research on mirror neurons has important features in common with his research on imitation.

The idea of a supramodal representation of action that we used to explain early imitation fits well with modern neuroscience discoveries about shared neural circuits for perception and action and so-called neural mirroring systems. (Meltzoff 2010, p. 19)

There have been a lot of research on brain reactions that seem to show that human beings’ and monkeys’ brains react in similar ways when the human being or monkey observe other people or other monkeys doing things as well as when they do these things themselves. My intention is not to deny these brain activities. Nor do I want to deny that we may often react spontaneously in a kind of imitative way to others. Still, there are some problems both with how some of the experimental research on mirror neurons is designed as well as with how some researchers interpret the results of the experiments.

One presupposition that appears to shape the empirical studies on mirror neurons is the idea that we generally understand other people’s actions by *observing* them. Generally the experimental studies in mirror neurons are constructed as situations where one person (or a monkey) *observes* another person (or monkey) who does something.

³¹ Rizzolatti and Arbib refer here to Donald (1991).

Then it is noted that certain similar events occur in both monkey's brains. (The whole concept of mirror neurons also metaphorically rests on something we mainly observe but do not engage with, i.e. mirrors.) Even when it comes to studies on communication these studies have the form of one person passively listening to another person uttering certain words, rather than the two persons engaging in a conversation together.³² The mirror neuron studies are generally not constructed as situations where two persons (or apes) are entwined in some form of *reciprocal* engagement like playing football, or two persons chatting or quarrelling with each other or two apes scratching each other's back. There seem again to be several reasons for why the empirical studies are built up in such a one-directional manner. One reason may be the earlier discussed influence from classical natural scientific research methods. Another reason may be that the studies suffer from similar problematic assumptions about interpersonal understanding as Meltzoff's studies on infants' imitative responses. This also means that Rizzolatti et al.'s theory on mirror neurons cannot be taken to merely lie on a deeper cognitive level than Bowlby's interactive dialogical conception of understanding. Rather it is based on a quite different conception of interpersonal understanding which is reflected in different kinds of examples of situations of interpersonal understanding as

³² This can for instance be seen in Rizzolatti and Craighero's article "The Mirror-Neuron System" (2004).

well as reflected in differences in the empirical research methods.

Would the theory on mirror neurons then become more tenable if a more interactive approach was taken in the empirical studies? It might become more tenable, but it is not certain. This depends on what the researchers claim that the brain activities in question mean. The theory on mirror neurons is influenced by a conception of understanding as private inner events. In such cases it is not evident that the theory would become more tenable even if a more interactive approach was taken in the empirical research. Even if one person "mirrors" another in some sort of interactive sense it is still not clear in what sense such a response would be expressive of understanding. Though I might have a similar brain reaction as you have, this does not in itself mean that I understand you. The concept of understanding is integral to the fact that we can also respond to each other in *different* ways, and it is integral to the meaning of the situation. It is integral to that we can consider a person's understanding as correct or incorrect, that we can criticize, console, help or quarrel with the person etc. In this sense my understanding of another person cannot be thought of as mere similar brain states that occur in two person's brains. By this I do not mean to say that it would be unimportant to study brain activities in connection with human understanding. Nor do I mean that it necessarily would be unimportant to study how one person's (or monkey's) brain reacts when the person (or monkey) observes another person (or monkey) doing something.

Clearly it can be an important thing to study. However, I am claiming that some of the mirror neuron theorists draw too far reaching conclusions about the meaning of such brain reactions. These conclusions are expressive of certain misconceptions of interpersonal understanding. I am also claiming that the empirical studies on mirror neurons tend to be constructed in a manner that appears to be influenced by a classical natural scientific research practice. This construction of the empirical research fits well with an epistemological, and analogical, conception of interpersonal understanding.

2.8 Sharing experiences

There is also a further problem with Rizzolatti et al.'s theory on mirror neurons as well as with Meltzoff's theory of imitation. Both rest on the idea that in order for us to understand another person we need to have the same inner impression or the same sensation as the other. Both theories also suggest that it is important that we can *share* another person's experiences. I shall reflect some more on what we mean when we talk of two people sharing an experience.

As I have noted earlier, Meltzoff suggests that imitation games can "deepen a sense of relationship". But then he explains this as being important because it makes the infant able to *mirror* its own behaviour with other people's behaviour. Thereby infants "gain a better sense of what their own felt acts look like". Surely infants can sometimes mirror their own behaviour with other people's behaviour,

and surely infants can sometimes gain a sense of their own behaviour through such mirroring. But is this the essential meaning of shared play? There are, I think, several aspects with what it means to share an experience that Meltzoff does not acknowledge. Consider the following description by Rudolph Schaffer (1977) of how a parent and child play together:

Watch a mother with her one-year-old sitting on her knee in front of a collection of toys: a large part of her time is devoted to such quietly facilitative and scene-setting activities as holding a toy that seems to be pushed out of range, clearing away those things that are not at present being used in order to provide the child with a sharper focus for this main activity, putting things next to each other that she knows the child will enjoy combining (such as nesting beakers), turning toys so that they become more easily grasped, demonstrating their less obvious properties, and all along molding her body in such a way as to provide maximal physical support and access to the play material. (Schaffer 1977, p. 73)

Schaffer describes here a situation where a parent and child play together. However, even if they share a moment together they do not do the same thing, nor is it evident that they feel the same thing. It might, for instance, well be the case that the parent is bored while the child is having fun. The child is not imitating its parent, nor is the parent imitating the child. Rather the shared play consists in that the parent helps the child along in doing things. But while

doing so they are also spending time together. The parent's helping creates room for the possibility to share a moment. Consider also the following description by Olli Lagerspetz (2008):

In what sense does a train ride from Sorrento to Pompeii constitute a shared experience? —If my daughter is riding the local train with me, we will receive almost the same visual and auditory input. Only she is shorter and seated a bit away. An unknown tourist next to me, being the same length, will receive stimuli that correspond to mine even more closely. But when I speak of the experience as something I share with my daughter, I do not think of it as something just accidentally shared, as it will be with the adjacent stranger. My experience is that of a trip *together with* my daughter, just as hers is a trip *with me*. (Lagerspetz 2008, p. 15)

Lagerspetz suggests, that when we talk about two people sharing an experience we say something about how the experience cannot be understood from an individual point of view. It is what the two *do together* that *is* the experience. This does, however, not have to mean that they feel the same thing. Lagerspetz suggests, by his example, that often what we mean by two persons sharing an experience cannot be understood by simply looking at the moment itself but is something that gets its meaning through the way the two persons also have shared a life history together. In this sense the sharing cannot be understood merely as two persons having similar feelings that are

“contagious”. Consider the following poetic reminiscence by P.F. Thomése of how he learned to see birds as a child. In his book *Shadow Child* (2005) Thomése writes:

One sees best through the eyes of another. ‘Look’, my father would say, and if I looked carefully I could see it too. We were outside, and what he pointed to in the bushes became birds. And the birds became different: chiffchaff, grosbeak, flycatcher. With my father’s eyes I saw them, each and every one.

Until he died. Suddenly they were gone, the trees were still, everything had lost its tongue.

Right after my father was gone, when he was, as it were, almost still alive, I had the feeling I had to act as his observer, in case his death proved a passing thing and he would have to be brought up to date afterwards.

Those were the days of an extreme keenness, because I was looking for two. It was as though I had to keep the world from falling apart, on my father’s behalf. (Thomése, 2005, p. 53)

Thomése talks about the way his father took him along to look at birds, and how the world became alive for him through this. He does not here describe two people imitating each other, and then somehow being affected by each other and feeling their similar feelings of joy. Thomése describes how their looking at birds *together* gave meaning to the whole situation. And it is also because this is a close relationship with a long history of a daily shared life that their looking at birds comes to have a special meaning as a part of this shared everyday life.

Thomése also talks about seeing in a way that is markedly different from the way philosophers usually talk about seeing. Usually seeing is considered to be an individual matter of observing the world of objects, where other people are merely a more complicated kind of object. But Thomése describes how he came to see the birds in the bushes through his father's eyes. The birds became alive, got form, got distinction for him through the way his father took him along to look at birds. This also suggests how the theory about mirror neurons as well as the theory on imitation are problematic. They are problematic because they presuppose that shared experiences mean that two persons have similar momentary inner states. Such a perspective does not allow for the way close personal relationships, with a long shared life history, give meaning to our ways of experiencing and doing things.

From the perspective of theory of mind (including simulation theory) there is also a tendency to emphasize a careful *attentive* attitude towards other people. Understanding is assumed to be an activity connected with such words as "attention", "reflection" and "concentration". However, in his book *The Absent Body* (1990) Drew Leder argues that the emphasis on attention reflects an old rationalistic Cartesian perspective on human understanding. But it also has to do with the idea that other people's intentions are *difficult* to discern. However, Leder suggests that our ability to relax and to be unaware of things are as essential expressions of understanding as our ability to be aware of things. The way children grow into various ways of being with others largely consists in

their growing into a *non-reflective* and *relaxed* attitude to others. This is not a matter of the child becoming inattentive. Rather it is a central and normal part of how we grow into close relationships that we can relax in each other's company. For instance, young children will commonly tend to *fall asleep* in the arms of others. A child's whole ability to fall asleep and her sense of becoming sleepy is usually shaped through the comforting presence of another. Tiredness can of course be a physical state and if you are exhausted you fall asleep anywhere. But being tired generally makes a child hysterical or angry or irritated. When it comes to children; relaxing and falling asleep is thus often dependent on others. As a child grows up she has to *learn* to sleep alone. A child's ability to sleep is in this sense often dependent on, and formed by, other people *acknowledging* her need to sleep. This kind of ability to relax in the bodily presence of another is not discussed from the perspective of theory-of-mind theory, because it is not seen as anything that enhances our ability to "read minds". But very much of how a child grows into a life with others takes this form, including how the child learns to talk. By this I do not mean that children never fall asleep by themselves, nor do I mean that a child always feels comforted by the bodily presence of others. Children's ways of relaxing as well as their way of responding to others vary, and parents behave in various ways towards their children. Someone might therefore criticize me here by saying that my example is merely one empirical observation among others. My suggestion is, however, that this example is reflective of a broader pattern of how

human beings can come to have meaning for each other and respond to each other. That we can relax in each other's company is largely part of how we grow into a close relationship and it can be an expression of understanding.

Leder (1990) has the following example where a conversation grows out of a shared walk in the woods:

I am walking in the forest with a friend. As we stroll we point out various things to one another. The colour of the leaves, a passing bird, the changing of the seasons. I adjust to my friend's pace and she to mine. I find myself enjoying things more and in a different way than when I had come alone. We speak of other topics beside the scenery. Of politics, mutual friends, movies each has seen. But then we lapse into silent enjoyment of our surroundings. [...] As discussed, when I walk with my friend through the forest I am not self-conscious about my movements and gestures nor focused on hers. Our bodies stand in cotransparency, ecstatically involved with a shared world. The structure of bodily disappearance is modified but fundamentally preserved in this being-with-another.

However, I can easily imagine a situation that would give rise to explicit body thematization. For example:

While walking with my friend in the forest I notice her surreptitiously sneaking glances at me. I become aware that she thinks that something is wrong with me: that my words, gestures and comportment are those of a seriously unbalanced man. As I describe a movie, she seems not to be imagining it along with me

but focusing upon the strange way in which I talk. She apparently is looking for signs of my derangement. As I point out something in the forest, she seems struck by the outlandishness of my gestures, not by what I am gesturing towards. (Leder 1990, pp. 94-95)

In the first part of the example Leder describes how two persons relax in each other's company. He also describes how this makes the two experience their surroundings in new ways. Their walking together makes them take an interest in new things, enjoying the sound of birds etc. In the example one can also see how a sense for the other's expression changes depending on the character of the situation. We are not all the time looking attentively for certain changes in the other's voice or searching for patterns in the other's facial expressions. And that we are *not* doing so can be expressive of how we experience a meaning in each other's presence. However, as Leder suggests, we do also sometimes look carefully at others; this can be expressive of a negative attitude towards the other, or expressive of uncertainty in the relation or expressive of worry about the other. Even if we may often look at each other while talking we do not usually concentrate on each other's face or search for certain traits or gestures, but our listening and talking to each other may still have a character of sensitivity. Looking another in the eyes is often a part of emphasising one's words or sharing a thought with the other. When I listen to what you tell me I might rest my eyes in your eyes. This can be part of my concentrating on what you are saying. I might also look

you in the eyes as an expression of being sincere in my words. In this way sensitiveness to the other's expression is a constant part in how we talk, but this does not necessarily consist in our being intensely aware of the other's behaviour and expressions.

Leder's example also suggests that there is often a spontaneous and floating openness in conversation. We often change subjects depending on what the other person wants to talk about and depending on what she comes to think about. In *Totality and Infinity* (1969) Emmanuel Levinas writes: "Language is not enacted within a consciousness; it comes to me from the Other [...]" (Levinas 1969, p. 204) Sometimes this floating openness can grow into a quarrell that suddenly arises from nothing. This is a feature that often can be seen with siblings. Their way of quarrelling is often expressive both of how they can relax in each other's company as well as expressive of the open character of their conversations. But often the openness also takes the form of spontaneous considerateness that is reflected in how our speaking about different topics is not decided by either one of us but we move into other topics in an indeterminate way, letting the discussion flow from one thing to another depending on what the other person wants to talk about. This floating character of conversations is also often one form of how we can relax in each other's company. Talking is in that sense often connected with considerateness towards the other, a readiness to follow her, to respond to her words rather than necessarily only to get one's own thoughts through. But the considerateness need not take the form of constant

careful attention to what the other person is saying. It can, so to say, be integral to the conversation being relaxed, but it can also be integral to conversations sometimes being tense or strictly focused. And, of course, we are not always considerate in our conversations.

My reason for bringing up these examples of sharing experiences has been to suggest that the research on imitation and on mirror neurons rest on a problematically subjective and private conception of what it means to experience something. This is why there is the idea that two persons can only understand each other if they experience *similar* feelings, or if they imitate each other's behaviour. By the examples above I have suggested that what it means to experience something often cannot be understood without seeing how these experiences get their form and meaning through what the persons *do together* and then also through the life history the two persons share. In Lagerspetz', Schaffer's and Thomése's examples, mentioned earlier, one can also see that the way these parents share an experience with their child is expressive of a form of acknowledgement of the child. It is not because parent and child are necessarily very much alike or because they happen to have the same interests that they come to share each other's company. Rather, there is an acknowledgment of the child in the way the parent brings the child along to do things. However, our ways of doing things with our parents or with our children are not always experienced as positive moments of sharing. Often when parent and child go out to share experiences everything quickly goes to pieces and the shared moment

ends in quarrels and screaming, because the child does not put on her clothes fast enough or she does not want to walk the wonderful walk in the forest because it is boring and she is tired and she wants to play computer games, and the parent's plans of sharing a moment with her child goes to pieces and she starts to scream. "Bloody hell! I took you along so that we would do something fun together and all you do is whine about going home!" There are often efforts to acknowledge another person that for some reason or other fail. Perhaps these efforts fail because we have too big expectations, or perhaps we fail because our child simply is in a bad mood and she does not want to do anything at all today except scream that everything is boring and in between hit her brother. Or then these efforts fail because the parent slept badly at night and he is not able to keep up a good mood and he loses his temper even though he tries to do something fun with the children. These kinds of failures to share moments are an integral part of a shared life. They are not cognitive-epistemological failures to understand, but have meaning in a relational and moral sense.

2.9 Conclusion

There are two central aspects to the idea that infants learn to understand other people through imitation of facial expressions. On the one hand this idea rests on the conviction that human beings have a natural inclination to be social, to respond to other people. It is claimed that our sociality originates in a bodily responsiveness to others

rather than in a capacity for logical reasoning, as was argued by Jean Piaget. This emphasis on our natural bodily responsiveness to others is, I think, an important aspect of the research on imitation. The research on imitation and mirror neurons can be seen as taking a critical stance towards tendencies to consider the development of interpersonal understanding from a strictly rationalistic and then also non-bodily angle.

However, I have also suggested that the research on infant's imitation of facial expressions is reflective of a problematic conception of interpersonal understanding. The emphasis on imitation as the basic cognitive method or mechanism that enables the infant to understand others rests on the idea that we understand each other through analogical imagination. This idea is also reflected in Rizzolatti's theory on the function of mirror neurons. The idea that interpersonal understanding rests on a capacity for analogical imagination is reflective of the idea that the second-person perspective is dependent on a first-person perspective when it comes to interpersonal understanding. Further, Meltzoff's theory on imitation reflects a body-mind dualistic perspective on human beings. I have also suggested that this assumption also partly shapes the empirical research in problematic ways. These are assumptions that also partly shape the research on mirror neurons.

The assumption that interpersonal understanding consists in one kind of cognitive function makes the researchers focus too much on imitation while they ignore all other kinds of interpersonal responsiveness that can

take place between parent and child. The cognitive conception of interpersonal understanding also makes the researchers construct a kind of one-directional test situations. Further, I argued, with a reference to Wittgenstein, Hertzberg and Cockburn, that the focus on infants' imitation of facial expressions portrays facial expressions as having a problematically rigid character. I also argued that Meltzoff's conception of interpersonal understanding rests on body-mind dualistic assumptions. These assumptions also shape the empirical research.

I then discussed John Bowlby's examples of the mutual responsiveness between a parent and child, a responsiveness that takes many forms (both positive and negative) which often do not consist in imitation. I also suggested that one cannot understand what it means to be present to a young child if one considers only the face. The ways we are present to a young child take a reciprocal form, and involves our bodily way of being in various ways. But our presence also gains meaning through the fact that we come to share a long life history where conversations will be an essential part of this shared life. I have also suggested that what it means for a child to grow into an understanding of others must be seen as dependent on the fact that we acknowledge the child. The concept of imitation will not reveal what it means to acknowledge another human being.

Chapter 3: Autism and theory of mind

3.1 Introduction

The theme of this chapter is the syndrome of autism and research suggesting that persons with autism lack a theory-of-mind function. According to proponents of theory of mind, people have a mental function that enable them to interpret other people's behaviour and thus eventually to see other people as minded, as having intentions and beliefs. It is, however, suggested by theory-of-mind proponents that persons with autism lack such a mental function to see other people as mental beings.

In core cases autism is a pervasive developmental disorder that affects a person's life extensively. The most striking and central feature in autism is the lack of social responsiveness. Children with autism are largely unresponsive to other people, and often have serious deficits in language. They generally do not seem to acknowledge the presence of other people, do not respond to other people's talk, and generally do not play with other children etc. Autism is, however, not a unitary phenomenon but rather a large spectrum of patterns. Because of this it is defined as "autism spectrum disorder" or ASD. Often autism is connected with intellectual disability but sometimes it is not. Children with autism can also have various physical problems. It is, however, the social unresponsiveness that is the most striking feature in autism, and it is this that researchers in theory of mind are interested in explaining.

My intention in this chapter is twofold. One aim is to discuss the claim that persons with autism lack a theory-of-mind function or that they are “mindblind”. A broader aim, which connects with the other chapters in this thesis, is to discuss the relation between empirical research and theory-of-mind theory.³³

³³ In the *Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition* (DSM-5), diagnostic criteria of autism spectrum disorder are described in the following way:

“A. Persistent deficits in social communication and social interaction across multiple contexts, as manifested by the following, currently or by history (examples are illustrative, not exhaustive [...]):

1. Deficits in social-emotional reciprocity, ranging, for example, from abnormal social approach and failure of normal back-and-forth conversation; to reduced sharing of interests, emotions, or affect; to failure to initiate or respond to social interactions.

2. Deficits in nonverbal communicative behaviors used for social interaction, ranging, for example, from poorly integrated verbal and nonverbal communication; to abnormalities in eye contact and body language or deficits in understanding and use of gestures; to a total lack of facial expressions and nonverbal communication.

3. Deficits in developing, maintaining, and understanding relationships, ranging, for example, from difficulties adjusting behavior to suit various social contexts; to difficulties in sharing imaginative play or in making friends; to absence of interest in peers. [...]

B. Restricted, repetitive patterns of behavior, interests, or activities, as manifested by at least two of the following, currently or by history (examples are illustrative, not exhaustive [...]):

1. Stereotyped or repetitive motor movements, use of objects, or speech (e.g. simple motor stereotypies, lining up of toys or flipping objects, echolalia, idiosyncratic phrases).

2. Insistence on sameness, inflexible adherence to routines, or ritualized patterns of verbal or nonverbal behavior (e.g. extreme distress at small

3.2 Autism as “mindblindness”

The psychiatrist Leo Kanner was one of the first to describe case studies of autism. In the article “Autistic disturbances of affective contact” (1943) he gives the following description of a five-year old boy with autism:

He paid no attention to persons around him. When taken into a room, he completely disregarded the people and instantly went for objects, preferably those that could be spun. Commands or actions that could not possibly be disregarded were resented as

changes, difficulties with transitions, rigid thinking patterns, greeting rituals, need to take same route or eat same food every day).

3. Highly restricted, fixated interests that are abnormal in intensity or focus (e.g., strong attachment to or preoccupation with unusual objects, excessively circumscribed or perseverative interests).

4. Hyper- or hyporeactivity to sensory input or unusual interest in sensory aspects of the environment (e.g., apparent indifference to pain/temperature, adverse response to specific sounds or textures, excessive smelling or touching of objects, visual fascination with lights or movement). [...]

C. Symptoms must be present in the early developmental period (but may not become fully manifest until social demands exceed limited capacities, or may be masked by learned strategies in later life).

D. Symptoms cause clinically significant impairment in social, occupational, or other important areas of current functioning.

E. These disturbances are not better explained by intellectual disability (intellectual developmental disorder) or global developmental delay. Intellectual disability and autism spectrum disorder frequently co-occur; to make comorbid diagnoses of autism spectrum disorder and intellectual disability, social communication should be below that expected for general developmental level. (American Psychiatric Association 2013, pp. 50-51)

unwelcome intrusions. But he was never angry at the interfering *person*. He angrily shoved away the *hand* that was in his way or the *foot* that stepped on one of his blocks [...] Once the obstacle was removed, he forgot the whole affair. He gave no heed to the presence of other children [...] (Kanner 1943, p. 220)

A bit further Kanner gives another description of a six-year old boy with autism:

He was led into the psychiatrist's office by a nurse, who left the room immediately afterward. His facial expression was tense, somewhat apprehensive, and gave the impression of intelligence. He wandered aimlessly about for a few moments, showing no sign of awareness of the three adults present. He then sat down on the couch, ejaculating unintelligible sounds, and then abruptly lay down, wearing throughout a dreamy-like smile. When he responded to questions or commands at all, he did so by repeating them echolalia fashion. The most striking feature in his behavior was the difference in his reactions to objects and to people. Objects absorbed him easily and he showed good attention and perseverance in playing with them. He seemed to regard people as unwelcome intruders to whom he paid as little attention as they would permit. When forced to respond, he did so briefly and returned to his absorption in things. When a hand was held out before him so that he could not possibly ignore it, he played with it briefly as if it were a detached object. (Kanner 1943, p. 224)

Kanner has many similar, detailed descriptions of children with autism. The descriptions are striking in the way the children seem to be largely unaware of other people. It is claimed by theory-of-mind theorists that these problems all derive from the lack of a theory-of-mind function.

Among the first to suggest that autism is due to a theory-of-mind deficit were the psychologists Simon Baron-Cohen, Alan M. Leslie and Uta Frith, in their paper "Does the Autistic Child have a 'Theory of Mind'?" (1985). According to them, the social and linguistic deficits that persons with autism suffer from cannot be explained as the result of mental retardation since there are persons with autism who are not mentally retarded. They also point out that mental retardation often does not lead to any social impairment. For instance, persons with Down's syndrome are generally highly social. Because of this they argue that the social impairments in autism must be due to an "underlying cognitive mechanism independent of IQ" (Baron-Cohen, Leslie and Frith 1985, p. 38) This cognitive mechanism they describe as:

[...] a mechanism which underlies a crucial aspect of social skills, namely being able to conceive of mental states: that is, knowing that other people know, want, feel, or believe things; in short, having what Premack and Woodruff (1978) termed a 'theory of mind'. (Baron-Cohen, Leslie and Frith 1985, p. 38)

In *Mindblindness, an Essay on Autism and Theory of Mind* (1997) Baron-Cohen writes:

Imagine what your world would be like if you were aware of physical things but were blind to the existence of mental things. I mean, of course, blind to things like thoughts, beliefs, knowledge, desires, and intentions, which for most of us self-evidently underlie behavior. (Baron-Cohen 1997, p. 1)

Baron-Cohen continues:

It is probably impossible to imagine what it is like to be mindblind, in the same way as it is impossible to imagine what it is like to be a bat³⁴. To live in a bat's world, in which objects are known by echo location, must impart a notion of objects so radically different from the notion that we obtain through vision that it may be beyond our imagination. Conversely, it is probably impossible for a mindblind person to imagine what it is like to be a mindreader. In the words of Sperber (1993), 'attribution of mental states is to humans as echolocation is to the bat.' It is our natural way of understanding the social environment.

The gulf between mindreaders and the mindblind must be vast. (Baron-Cohen 1997, p. 4)

A few pages later Baron-Cohen continues:

Tragically, mindblindness is not an idle thought experiment or a piece of science fiction. For some people, it is very real. [...] In this book I will discuss the idea that children and adults with the biological

³⁴ Baron-Cohen refers here to Nagel (1974).

condition of autism suffer, to varying degrees, from mindblindness. (Baron-Cohen 1997, p. 5)

Baron-Cohen maintains that the syndrome of autism can be defined as mindblindness, or more precisely as a dysfunction of a theory-of-mind mechanism. Baron-Cohen and John Swettenham (1997) write:

[...] the theory-of-mind deficit in the majority of cases with autism is very severe. It has the potential to explain the social, communicative, and imaginative abnormalities that are diagnostic of the condition [...] (Baron-Cohen and Swettenham 1997, p. 884)

If one looks at real life descriptions of autism, such as those of Leo Kanner, Baron-Cohen's characterisation of autism as mindblindness can seem fitting. However, there are also other kinds of descriptions of autism that give a somewhat different picture of the condition.

In her book *The Siege* (1967) Clara Claiborne Park writes about her daughter who was born with autism. She begins the book with the following description:

WE START with an image—a tiny, golden child on hands and knees, circling round a spot on the floor in mysterious, self-absorbed delight. She does not look up, though she is smiling and laughing; she does not call our attention to the mysterious object of her pleasure. She does not see us at all. She and the spot are all there is, and though she is eighteen months old, an age for touching tasting, pointing, pushing, exploring, she is doing none of these. She does not walk, or crawl up

stairs, or pull herself to her feet to reach for objects. She doesn't *want* any objects. Instead she circles her spot. Or she sits, a long chain in her hand, snaking it up and down, up and down, watching it coil and uncoil, for twenty minutes, half an hour—until someone comes, moves her or feeds her or gives her another toy, or perhaps a book.

We are a bookish family. She too likes books. Rapidly, expertly, decisively, she flips the pages, one by one by one. Bright pictures or text are the same to her; one could not say that she doesn't see them, or that she does. Rapidly, with uninterrupted rhythm, the pages turn.

One speaks to her loudly or softly. There is no response. She is deaf perhaps. That would explain a lot of things—her total inattention to simple commands and requests, which we thought stubbornness; the fact that as month follows month she speaks no more than one word or two, and these only once or twice in a week; even, perhaps, her self-absorption. But we do not really think she is deaf. She turns when you least expect it, at a sudden noise. The soft whirr as the water enters the washing machine makes her wheel round. And there are the words. If she were deaf there would be no words. But out of nowhere they appear. And into nowhere they disappear; each new word displaces its predecessor. (Park 1967, pp. 3-4)

This was the state of Clara Park's daughter Elly³⁵ when she was a baby and to various degrees up to the age of four.

³⁵ Elly's real name is Jessica, but in order to make the discussion easier in this text I will use the name Elly as it is used in Park's book. In *Exiting*

Her daughter was a happy child, often in a good mood, but she was not active in the way normal children are. She did not try to explore her surroundings. She did not try to climb up stairs, did not try to get things. She was basically sitting or circling a spot. Nor did she generally seem to notice the presence of others. It was as if she did not see or hear, nor did she respond to others. A bit further in the book Park writes:

Elly did not point. Nor did she try to get objects that were not within her reach; she seemed unconscious they were there. Content in crib or pen, when removed from them she crawled freely from room to room. But it was motion, not exploration. She did not push or poke, open drawers, pull at lamps or tables. At twelve months, when she began to crawl, I got ready the gates that we had used to keep the other three children from falling downstairs. I never used them. Elly did not try to go down, and there was never a question of her falling. Unconscious of so much, she was conscious of the location of every edge or limit; she could be left safely on any bed. (Park 1967, p. 7)

Park describes how her daughter did not engage in doing things; she was to a large degree inactive. But at the same time her inactivity did not have to do with any kind of physical inability to move or clumsiness or complete

Nirvana: a Daughter's Life with Autism (2001), a later book by Park about her daughter, Park uses her daughter's real name, Jessica. I will use the name Jessica when I quote passages from this later book.

unawareness of her surroundings. She was on a certain level aware of physical limits.

Park also tells of there being occasions when Elly did react both to sounds and objects and also to people. Elly was also occasionally fast at picking up how to do things. But even though she for a moment seemed fully to know how to do something, these newly learned skills often evaporated into the air.

She made no move to climb the stairs, but one day, in play, her sister taught her to crawl up. She learned easily enough, and I thought the new skill would mean the usual extension of a baby's possibilities. She learned on a Friday. We went away that weekend to a stairless household, and when we came back the Sunday it did not occur to us to review her new skill. It was six months before she crawled up again. (Park 1967, p. 7)

There are a large number of such descriptions of Elly learning something and then not doing it again for months or even for years. She learned certain words but suddenly completely stopped using these words. She learned certain simple games but suddenly did not play them at all anymore.

From the earlier quote by Kanner one can get the impression that autistic children have no trouble in understanding or seeing objects while they are unaware of other people. However, from Park's description it appears as if her daughter was unaware of almost anything, i.e. *both* objects *and* people. How should this difference in

description be understood? One way to explain the difference is that children with autism are not all alike. Another explanation may be that the children Kanner describes are several years older than Park's daughter Elly. It might well be that the children in Kanner's description had been in a similar state as Elly was, when they were younger. Perhaps they had developed and become more active and aware of objects. Another aspect to note is that Kanner's description of how the children play is quite short. He says in one of the quotes that the boy "showed good attention and perseverance in playing". It is, however, common that children with autism are interested in objects in abnormal ways and then also "play" in abnormal ways. For instance, children with autism often like to spin things endlessly. Kanner also notes in the first quote that the boy preferred to play with objects that could be spun. In Park's description Elly flips pages in books endlessly. Such behaviour is not expressive of a normal sense for objects.

Park describes several traits in Elly's behaviour, traits that involve *both* her failure to take an interest in objects *and also* her failure to acknowledge or respond to people. Park also describes a changing character in Elly's sense for her surrounding; she does not always appear unresponsive to people or to her physical surroundings. There is a general lack of spontaneous engagement in, or responsiveness to her surroundings, but at the same time Park also tells about Elly being sensitive when moving around in the house and sensitive to certain sounds. There is also the general pattern of Elly not responding to other people, while on the other hand every morning when

waking up Elly happily greets her mother. She also loves to be tickled by her father. Jenny Lexhed (2008) also writes of how her autistic son loved to be tickled and how during these tickling sessions he used to laugh and look intensely into his mother's eyes. There are also the sudden displays of new skills such as when Elly's sister one day taught Elly to climb up the stairs, a skill that also then equally suddenly disappeared. Thus there are various patterns of apparent incapacity or lack of engagement but also evident displays of sudden capacity and development and learning combined with an equally sudden incapacity or lack of interest in continuing doing things. And there are various patterns of unresponsiveness to others combined with certain patterns of evident responsiveness and presence to others.

My suggestion so far is that theory-of-mind theory is influenced by certain careful and very striking real life descriptions of autism, such as Leo Kanner's. Even if Kanner's descriptions are carefully made one can be struck by the apparent difference in how autistic children seem to be unaware of people while they seem to be aware of objects. From this one can get the impression that autism essentially consists in "mindblindness".³⁶ However, I have

³⁶ I do not mean to imply that Kanner's descriptions of children with autism are not good descriptions. However, there are certain differences in the character of these descriptions. Kanner describes children that are several years old. His descriptions are also to a large extent focused on the child's behaviour during the child's visit at the clinic, though he does also describe the children's life history briefly as well as their further development. Park's description of her daughter

also suggested that there are other descriptions of autism that do not fit as well into the theory of “mindblindness”. In Park’s description of Elly as an infant there is not a stark difference between Elly’s sense for objects and her sense for people. Since Elly, as a young child, appears to be largely unaware of *both* objects and people it is not clear that Elly’s state could be defined as “mindblindness”. But also the fact that Elly occasionally does respond to other people, for instance that she loves to be tickled, goes against the theory of mindblindness. The claim that autism can be defined as “mindblindness” rests on the idea that we understand other people by the use of a mindreading mechanism. Even though there are many features in autism that can seem to support the theory of mindblindness I will argue that autism cannot be understood from such a perspective on interpersonal understanding.³⁷

stretches over a longer period of time. She describes her daughter’s behaviour from infancy to adulthood. She also describes her daughter’s behaviour in ordinary life situations. Because of this difference in focus, I think Park’s description of her daughter reveals aspects that are not as clearly discernible in Kanner’s descriptions.

³⁷ The temptation to describe autism as a kind of body-mind dualism, where the person with autism only can see outer bodies but not minds, is also reflected in the following imaginary description of autism, by Alison Gopnik: “This is what it’s like to sit around the dinner table. At the top of my field of vision is a blurry edge of nose, in front are waving hands.... Around me bags of skin are draped over chairs, and stuffed into pieces of cloth, they shift and protrude in unexpected ways.... Two dark spots near the top of them swivel restlessly back and forth. A hole beneath the spots fills with food and from it comes a stream of noises.

3.3 Autism, pretend play and mindreading

In discussions concerning autism as mindblindness, it is often concluded that pretend play is an important practice whereby a child learns to imagine that other people have minds. The philosopher A.I. Goldman writes in *Simulating Minds: The Philosophy, Psychology and Neuroscience of Mindreading* (2006):

[...] creative role play constitutes a kind of simulation [...] such practice at simulation makes a positive contribution to children's mindreading. [...] Normal children engage in role play from the age of 2, acting out the role of a person or creature. Role play [...] is a species of pretend play in which a child impersonates a character, such as a mother, a bus driver, or a soldier. One can also project a role onto an object like a doll or toy, which serves as a prop for the role. I shall interpret role play as *extended* imitation. [...] role play is

Imagine that the noisy skin-bags suddenly moved toward you, and their noises grew loud, and you had no idea why, no way of explaining them or predicting what they would do next." (Gopnik 1993, taken from Baron-Cohen 1997, pp. 4-5) Gopnik's description of autism is an imagined case. It is not a description of a real case of a person with autism. One can, however, see in the quote how the theoretical idea of "mindblindness" is reflected in a tendency to give a strongly dramatized body-mind dualistic description of the autistic person's impression of people. Baron-Cohen suggests that Gopnik's description is an accurate description of autism. I think Gopnik's description is dramatised in a problematic sense. Such dramatised imaginative descriptions will not help people to understand persons with autism. It will, on the contrary, make people estranged to persons with autism.

'extended' imitation insofar as it involves more than mere *behavioral* copying. It involves *imitation* as well, that is, attempts to enact in one's own mind a target's mental states or processes. Evidently children's role play involves such mental simulation, as manifested by verbal and nonverbal behavior. (Goldman 2006, p. 196)

According to Goldman, children's capacity for role playing and playing with dolls is central for the development of interpersonal understanding because it enhances a form of analogical imagination whereby the child learns to understand other minds. Autistic children's inability to use such an analogical form of imagination is, according to Goldman, one reason why they do not develop an understanding of others. Baron-Cohen writes:

If children with autism really have some impairment in the development or functioning of ToMM [Theory of Mind Mechanism], they should also have difficulty understanding the mental state of pretending. [...] A range of studies now show that in children with autism spontaneous pretend play is severely impoverished or altogether absent. (Baron-Cohen 1997, pp. 76-77)

The philosopher Gregory Currie also concludes about autistic persons: "We have seen that autistic individuals show coincident deficits on pretend play and on the comprehension of the mental states of others." (Currie 1995, p. 159) These deficits in pretend play are considered to reflect the fact that children with autism lack a capacity for

analogical imagination which in turn is considered important for the theory-of-mind mechanism to work.

Children who play with dolls often imitate various situations in real life. In this sense children's play often takes on a kind of analogical character. The question is, how should this analogical character of playing be understood? Does the abundance of forms of playing that involve a kind of analogical imagination, such as for instance doll play, show that the social importance of playing with dolls lies in the ability to use a method of analogical imagination that enhances the child's capacity to imagine the mental states of other people? Is children's pretend play expressive of a training for mindreading? Does the lack of pretend play among children with autism point towards a theory-of-mind deficit?

There are several problems here in how pretend play is taken to enhance the child's capacity to read minds. One problem is that pretend play is portrayed as if it were basically a lonely child playing with her doll. However, playing is centrally something children do with *others*; it largely has meaning as a social way of being with others. If a child were *only* interested in pretend playing by herself and not at all interested in playing *with* others, there would be something seriously wrong with this child. Indeed children often want to play by themselves, but equally often they want to play *with* others. The emphasis on pretend play in theory-of-mind theory is connected with a consistent tendency to ignore the fact that children play with others. This is reflective of the way the proponents of theory of mind think that to understand another person

means to learn to read the person's intentions rather than to learn to do various things together with the other person.

Another problem with the idea that pretend play enhances the child's development of a theory-of-mind function is that only *one* form of playing is emphasised while a lot of other forms of playing are totally ignored. Pretend play is merely one form of playing that is part of a whole lot of *other* ways of playing, and also part of a lot of other ways of being together, that have nothing to do with analogical imagination or with imitation. Children play hide and seek together, they fight with each other, parents and children eat together, they go along shopping, they chat about various things almost constantly, they quarrel etc. I do not mean to say that pretend play is unimportant for how a child grows up and develops an understanding of various things, but it is problematic to suggest that pretend play is a basic tool by which the child learns to read minds. In playing, as much as in our other forms of being together, we are confronted with each other in ways where we learn to be considerate as well as mean, we learn to accept the wishes of others, we learn to help each other, and we learn to control our temper. But we also learn mean tricks of ostracising someone, not taking the other kid along to play etc. We get used to people having their various good and problematic attitudes, and we get used to quarrelling and getting irritated with each other and expressing our opinion or keeping quiet. However, from the perspective that pretend play and imagination is connected with learning to read other minds, none of these

social and moral aspects are considered as expressions of how a child learns to understand others.

My argument so far has been that theory-of-mind theory portrays playing in a problematic one-sided and restricted light, as if it were a tool by which the child learns to read other minds. If one thinks of playing as a tool for mindreading one will miss how crucial it can be for a child with autism to learn to play. I shall now again return to take a look at Park's description of how she tried to get contact with her daughter Elly:

Elly is prone on the floor, legs frogged out on either side of her. She is under a blanket and so invisible, but I know the position and the steady rhythm that goes with it. She is rapt, removed, she needs me not at all. I crouch beside her, ready to enter her world in a way she can appreciate if she will and ignore if she wants to. My finger goes under the blanket, then my hand. No response. My head follows. Elly knows I am there. There are two of us now, withdrawn from the world but near each other. It is very inward, warm and dark—a physical expression for undemanding intimacy. There is nothing difficult here—nothing to do, nothing to say. The only thing you need is time and the willingness to spend a lot of it with your head under a blanket.

It became possible to make a game out of raising the edge of the blanket. By the time she was two we could get her to play the peek-a-boo game whose absence we had noticed at ten months, and she even made a little 'there-she-is' noise to go with it. As time went on we moved forward, but not far. She began to

welcome me into her enclosures. At three and a half she even developed on her own a new discovery game—herself closed in a closet, I to open the door. Better yet, both of us sit quiet in the dark closet, door pulled to, she and I, close together, everything else shut out. We still do that sometimes, even today. (Park 1967, pp. 92-93)

Here Park gains contact with her daughter by simply lying beside her under a blanket for long times. Slowly out of this they develop a peek-a-boo game. And then Elly herself develops the game further. No examples like these are discussed by Baron-Cohen. From the perspective of theory-of-mind theory, bodily closeness plays no role, even though bodily closeness is one of the most fundamentally important ways of having contact with a young child. Another such attempt to get Elly to respond has to do with playing with toes.

How could we give Elly's shapes a human meaning? Elly was three years old, and I was still trying to find out whether she recognized that a doll had a human shape. Sculpture, which reaches touch as well as sight, is one degree less abstract than painting. We sit on the floor with a small girl doll. It belongs to Elly's sisters; it has many outfits. I dress it. Elly pulls the clothes off, chooses another costume, we begin again. The game holds her interest over several weeks. Can I assume that it shows she knows the doll represents a human body? Testing, testing. In the absence of other evidence I cannot be sure. I try to put the doll into interesting situations, but of course they are not interesting to Elly.

One day, however, an idea floats into my mind which most of the time is vacant; I play 'this little piggy,' which Elly knows, not on Elly's toes but on the doll's. Elly shows no interest, but in her bath that night I surprise her counting over the doll's toes, ending with the delighted squeal that for her signals the climax of a tickling game. It seems unmistakable that she is tickling the doll, that it is safe to conclude at last that she sees the doll has toes like her own.

I can make explicit now, the principle that I then perceived so dimly that I made use of it only by accident: in reaching the eyes and ears of such children, and later on their minds, one must begin with sensations their bodies can recognize. From Elly's toes to the doll's. It is not for three full months that it occurs to me, as mechanically we turn the pages of *A Treasury of Art Masterpieces* (so much more interesting for mother than *Little Golden Books*), to play 'this little piggy' on the bare toes of those Renaissance Christ-babies. Which I do. And Elly *laughs*. (Park 1967, p. 61)

This case can appear to resemble Baron-Cohen's and Goldman's claim that doll play and pretend play are essential in order for a child to learn to understand other minds. However, Goldman does not at all talk about pretend play as something we do with *others*. Neither does he say anything about how children can enjoy a kind of bodily playfulness, such as tickling. For Goldman doll play is important through enhancing an analogical form of imagination. However, for Elly the picture of the Christ-baby became meaningful through the way Elly had enjoyed it when her mother played this-little-piggy with

Elly's toes. This resembles the situation of their both lying together under the blanket and it slowly developing into the peek-a-boo game.

In his book *Autism and the Development of Mind* (1993), the psychiatrist Peter Hobson writes:

If personal relatedness involves intersubjective exchanges that are co-ordinated between the infant and others [...] then we shall need to think in terms of the structure of *interpersonal* events. Even if one wishes to maintain a focus on the individual infant, it is still necessary to consider how specifically interpersonal patternings of behaviour and experience are generated and registered. [...] More specifically, the argument goes, a child's experience of affectively patterned personal relatedness is constitutive of the child's understanding of the nature of persons with minds. (Hobson 1993, p. 186)

Hobson's conception of interpersonal understanding differs from the conception put forward by theory-of-mind theorists. According to Hobson, autism ought to be seen as involving various difficulties of *engaging* with others. A child's ability to grow into an understanding of other people is largely dependent on the way the child is entwined in affectionate relationships with others where the child responds to these others spontaneously and these others respond to the child.

[...] there is no radical developmental disjunction between the perception of 'bodies' or 'behaviour', and the apprehension of 'mind'. *A fortiori*, it is not a matter

of the infant beginning with the cool perception of thing-like bodies, and only subsequently interpreting or theorising that behind bodily behaviour there might be 'mind'. On the contrary, aspects of mind are perceived in aspects of expressive behaviour. [...] To perceive personal meanings is also to have the propensity to react to such meanings in appropriate ways. Perception is relational; in the early stages of personal relatedness, perceiving has intrinsic connections with feeling and acting. (Hobson 1993, p. 187)

Hobson emphasises how the child grows into interpersonal understanding through being part of affectionate relationships. Baron-Cohen's conviction that interpersonal understanding is *fundamentally* a matter of seeing that other people have an inner mental life makes him pay no attention to the meaning of close relationships. Consequently he pays no attention to descriptions of how parents of children with autism try to get contact with their child, and how these repeated efforts largely take an affectionate form. Hobson also emphasises the importance of sharing:

[...] it is partly from early experiences of 'sharing' [...] that a child's concept of person is derived. It is to bodily expressed attitudes that a young child responds, often with (sharing) attitudes of her own. (Hobson 1993, p. 188)

In his article "The Roots of Mindblindness" (2004) the philosopher Stuart Shanker writes:

[...] an infant's emotions develop in the context of co-regulated and shared emotional experiences (i.e. interactions in which partners are continuously active, continually adjusting their behaviours to each other), including the capacity to recognize the significance of or respond automatically to facial expressions of emotion [...] Far from being predetermined phenomena, the emotions that a child experiences and the capacity to understand the emotions that another person is experiencing are two aspects of one and the same phenomenon, both shaped by the nurturing relationships that a child experiences with her primary caregivers. (Shanker 2004, p. 692)

It is, according to Hobson and Shanker, largely difficulties on this level that a child with autism struggles with. However, it is also important to note that parents try in various ways to create forms of contact, as one can see in the quotes by Park above. There is so to say no absolute wall surrounding a child who has autism, even if it might be very difficult to establish contact, to arouse the child's interest in things and to figure out ways to engage the child in interpersonal relationships. Most importantly, the situations of playing described by Park above are *not* proof of the pretend play theory put forward in theory of mind, because according to the pretend play theory no shared relationships are of relevance at all for the meaning of playing. That is, children with autism do indeed often tend not to be good at pretend play, but this does not mean that pretend play enables a child to read minds. What a child

lacks when it is not good at pretend play is something other than a tool to learn to read minds.

However, there is also another side to the way Elly comes to recognize the picture of the baby through the playing with toys. It is something that can be seen in the following quote:

Elly becomes interested in her brother's kindergarten workbook. She turns the pages as always, looking with attention but without recognition. But now I have an inkling of how to proceed. As we pass a large, realistic picture of an ice cream cone, I take her hand and make her pat it. Next time she looks at the book there is a pause in the mechanical turning; that picture, at any rate, she sees.

This book was full of usable sights; for the picture of a school playground I made her fingers walk up the slide and go 'whee' down, I made them ride the seesaw and the swings. I no longer wondered about her comprehension; her delight left no doubt of it. (Park 1967, p. 62)

There is a difference between an ordinary child and Elly in that it was to such a large extent difficult for Elly to engage in doing things. For an ordinary child doing things, touching, pulling, pushing, etc., is central for her coming to see a meaning in things. This is entwined with parents and others also constantly showing, helping, taking along, forbidding, etc. But even though Elly's parents tried to engage her, Elly still largely lacked that capacity to become interested in doing things. At many instances Elly used her

mother's hands to do things. It was not merely as if she used her mother as a tool but rather as if she was unable to engage in doing things by herself; as if she needed another for this. She largely did not respond to invitations to take an interest in things by herself or by doing things herself. But when Elly reacted to the "this little piggy" game her mother realized that her daughter still, despite her general inactivity, sensed a meaning in things through touch. This was a very important insight since, contrary to ordinary children, Elly did not as easily seem to become aware of things through her sight or hearing. Therefore touch became a very important source for contact and understanding. It was through meaningful touch in the form of doing things as a kind of imaginary bodily playing or bodily presence (such as in lying together under the blanket) that Elly also came to *see* things with her eyes.³⁸

³⁸ Park also describes how Elly could communicate by singing even though Elly was to a large degree unable to utter specific words. There is, I think, something similar here with how she came to sense a meaning in pictures through bodily "acting" the picture. Park writes: "Tunes became words for Elly. 'Ring around a rosy' was the first. She was three and three quarters that spring, and she had been playing the game for many months. Now her new musical alertness picked up the tune. As soon as it did, she extended it spontaneously to a picture of children in a ring, then to a garland of flowers, and finally to the unadorned figure of a circle. The song—shortened to its first few notes—for more than a year remained her word for 'circle' and the cluster of ideas around it, functioning far more reliably than any of her actual words.

Other leitmotifs followed. 'Happy birthday' equalled cake and, by extension, candles and fire. [...] 'London Bridge' became a bridge motif; the dwarfs' song from *Snow White* did duty for 'dig.' We noticed that

Through this her mother saw a way to establish contact with her child, which would eventually enable them to communicate with each other by the use of pictures. Often the tendency among children with autism to use their parent's hands as if they were mere "tools" is taken to imply that the child is not aware of other people, treating them merely as objects. Surely this can be the case, but perhaps at least for some autistic children this behaviour might partly have to do with an inability to take an initiative on their own and partly with a difficulty of seeing things without touching the thing in a meaningful way? At least Elly's case seems to imply some such a reaction. Park describes several instances where it is as if Elly is unable to take initiative on her own even though she wants to do something. In the following description Elly is about three years old:

Though she is fascinated with water, activating a faucet is harder. [...] she learns of herself to use the kitchen faucet, which requires only a simple push. But an ordinary faucet requires both pressure and twist. I put Elly's hand on it; wrist and fingers go limp. My whole hand covers Elly's; using her as a tool, I turn on the faucet. This first time, and again, and again, all the force is mine. Elly likes water and she has no objection

though she now sang many songs freely, she never sang her leitmotifs at random or for their own sake as songs. Nor did she sing them musically, like the others, but rapidly, schematically, *functionally*—only just well enough for them to do their job of communication. Music was serving as her avenue to words [...]" (Park 1967, pp. 83-84)

to repetition. Imperceptibly—I hope it is imperceptibly—I lighten my pressure. The small hand beneath mine is no longer quite limp. It seems that there are muscles there after all. I move my hand a quarter inch up hers as I turn the water on again. Another quarter inch. A half. Infinitely gradually I withdraw my hand, up her fingers, up her arm. Finally all that is left is one finger on her shoulder, to enable her to maintain the fiction that it is I, not she, who is performing the action. [...] next day we must go over the process again to re-establish the skill, but we can do it more rapidly. Then I remove my finger; my presence now is enough. The next day she does it alone, thrilled, delighted, over and over. (Park 1967, pp. 51-52)

Hobson also reflects on the fact that autistic children do not play much with things. He notes that autistic children often do not show much spontaneous play while they do play when being elicited. He argues that the lack of spontaneous play often seems to have more to do with a lack of motivation than with cognitive deficits.

[...] there is the specially marked abnormality in spontaneously produced as opposed to elicited play. Why do so very few autistic children seem to have fun in playing? As Sharon Wulff (1985) describes, an autistic child who is left to his or her own devices in a playroom full of toys is very likely to ignore the toys and continue rocking or hand-flapping, or will spin moveable parts rather than becoming engaged in a meaningful way. (Hobson 1993, p. 164)

Hobson argues that the ways we grow up to see a meaning in things, to enjoy playing with things, doing things, handling things, moving about, is largely entwined with how we are responsive to others. Elly's inability to become spontaneously interested in, or to react to, her physical surrounding seems to have been part of her inability to engage with people.

In the living whole, nothing comes first. Work done on any one of Elly's inabilities affected the others. Every game we played, every exercise we devised to extend Elly's use of her body, her eyes, her ears, her voice, her mind, worked in addition to breach that jealously guarded isolation which for those who lived with her remained the most obvious and the most terrible aspect of her condition. (Park 1967, p. 88)

Park again:

But all this—head under the blanket, doll play, practice in imitation—was for one purpose: to bring her into contact with people. The evidences of progress were small, but they were beginning to accumulate. In the weeks before her third birthday, all these things happened. In the course of a tickling game, *she* poked *me* with her finger, to her great amusement. (It did not happen again for six months.) She fed me a candy, as she did a little later at Dr. Blank's, putting it into my mouth herself, not merely pushing my hand to do the work. When an elderly gentleman had held her hand and tickled it, she held out her hand to invite him to begin again. She even clowned a little for him, as a

normal baby would. One memorable afternoon she spontaneously hugged her sister. Three or four times she pushed the children, not with the detached don't-bother-me attitude we were used to, but with the first anger and hostility she had ever shown. [...] there was a general forward motion that helped us bear setbacks. As Elly approached four she abandoned doll play altogether and resisted all attempts to lure her back to it. But other things took its place. The new ability to joke and tease did not disappear. She spilled water on me on purpose, and laughed. She turned the light off while we sat at supper. Teasing is not an autistic activity.

We were able to establish a few reciprocal games—ones in which Elly too must play her part. Elly, who six months before would lackadaisically roll a ball back to you from twelve inches away, would now retrieve it with enthusiasm if you threw it several yards. Out on the wide college lawns, I could now do as I had delighted to do with the other children—crouch down and hold out my arms while a small, laughing creature came running from fifty feet away to end in my embrace. (Park 1967, pp. 102-103)

These responses are not preliminary steps of seeing an outer physical human shape towards a “real” understanding of other people as “mental”, they are genuine responses to another. It is an essential feature in human life that we grow into such various forms of spontaneous affectionate responsiveness to each other. And it is also through this that we grow into slowly having a more advanced sense of others, which for instance shows

in how we get used to respond to our close ones' attitudes of irritation or jokes or anger or pain or forgetfulness. However, no such descriptions of expressions of interpersonal responsiveness are mentioned when theory of mind is emphasized as the basic problem in autism.

3.4 Autism and the false belief task

Now someone could object that so far I have discussed autistic children that are on a much lower level of understanding than the cases that are generally discussed by theory-of-mind theorists, and that I hereby sidestep their question. Partly this critique is correct and partly not. It is not correct in the sense that by discussing the case of Elly as a young child I have not wanted to sidestep their question but I have wanted to argue that it is important to acknowledge the life history of children with autism, and then also important to acknowledge that children with autism develop. By discussing the case of Elly I have also wanted to question a certain kind of one-sided attention, among theory-of-mind proponents, to certain features in autism. This one-sided attention is reflective of the body-mind dualistic, rationalistic and epistemological conception of interpersonal understanding that theory-of-mind theory consists of.

However, there are features in older and more able children with autism that seem not as stark in younger or less developed children. Such features can concern how children with autism often fail to consider other people's perspective on a situation, or in other words fail to see that

other people have intentions. These features can appear to support the theory that children with autism specifically have a theory-of-mind deficit. Other features also show in the autistic person's use of speech. In what follows I shall therefore first discuss some studies concerning autistic children's incapacity to understand that other people have intentions or beliefs. After this I discuss some studies indicating that the language deficits that children with autism can have indicate a lack of theory of mind.

There is a well-known psychological test that concerns the ability to consider other people's perspective, or in other words the ability to see that other people have beliefs and intentions. This test is called "the false belief task". Originally the theory about false beliefs was put forth by Daniel Dennett (1978). In his article "Beliefs about Beliefs" Dennett reflects on whether chimpanzees might have a theory-of-mind function or not. He proposes that one might try to find out whether this is the case by making a false belief experiment. A few years later Heinz Wimmer and Josef Perner (1983) put Dennett's thoughts into action by constructing a real test situation. They conducted the false belief task with ordinary non-autistic children between the age of three and four. Since then the false belief task has been repeated a large number of times and in varying forms, with ordinary children at the age of three to four, as well as with autistic children. The idea has been that the children's capacity or incapacity to accomplish the test indicate that children gradually develop an

understanding of other people as minded. Here is a description of Wimmer and Perner's false belief task³⁹:

In order to test the subjects' comprehension of the other person's wrong belief, stories like the following were constructed: A story character, Maxi, puts chocolate into a cupboard x. In his absence his mother displaces the chocolate from x into cupboard y. Subjects have to indicate the box where Maxi will look for the chocolate when he returns. Only when they are able to represent Maxi's wrong belief ('Chocolate is in x') apart from what they themselves know to be the case ('Chocolate is in y') will they be able to point correctly to box x. This procedure tests whether subjects have an explicit and definite representation of the other's wrong belief. (Wimmer and Perner 1983, p. 106)

The result of this test was that a four year old child generally understood that the story character Maxi will have a false belief while a child of three generally did not realise this. The test seems to indicate that it is difficult for children under the age of four to complete the task because they have an undeveloped mental capacity to see that other people have intentions and beliefs.

A few years later (in 1985) Simon Baron-Cohen, Alan M. Leslie and Uta Frith built a similar kind of test that they conducted with autistic children. Their task was named the

³⁹ Wimmer and Perner made many variations on this false belief task. I have only quoted one of these. See Wimmer and Perner (1983).

Sally-Anne task. They describe the task in the following way:

There were two doll protagonists, Sally and Anne. First, we checked that the children knew which doll was which (Naming Question). Sally first placed a marble into her basket. Then she left the scene, and the marble was transferred by Anne and hidden in her box. Then, when Sally returned, the experimenter asked the critical Belief Question: 'Where will Sally look for her marble?'. If the children point to the previous location of the marble, then they pass the Belief Question by appreciating the doll's now false belief. If however, they point to the marble's current location, then they fail the question by not taking into account the doll's belief. These conclusions are warranted if two control questions are answered correctly: 'Where is the marble really?' (Reality Question); 'Where was the marble in the beginning?' (Memory Question). (Baron-Cohen, Leslie and Frith 1985, p. 41)

The result of their false belief task was the following:

23 out of 27 normal children, and 12 out of 14 Down's Syndrome children *passed* the Belief Question on both trials (85% and 86% respectively). By contrast, 16 of the 20 autistic children (80%) *failed* the Belief Question on both trials. [...] All 16 autistic children who failed pointed to where the marble really was, rather than to any of the other possible locations [...] (Baron-Cohen, Leslie and Frith 1985, p. 42)

As one can see from the results of the false belief task there was a striking difference in ordinary children and children with Down's Syndrome answering correctly while children with autism generally failed. Baron-Cohen, Leslie and Frith conclude:

Our results strongly support the hypothesis that autistic children as a group fail to employ a theory of mind. We wish to explain this failure as an inability to represent mental states. As a result of this the autistic subjects are unable to impute beliefs to others and are thus at a great disadvantage when having to predict the behaviour of other people. (Baron-Cohen, Leslie and Frith 1985, p. 42)

The results of the false belief task indicate that children with autism have fundamental problems with understanding how other people might think in certain situations, problems that ordinary children over the age of four, and children with Downs' syndrome, generally do not have. According to Baron-Cohen, Leslie and Frith, the result of the test indicates that autism has to do with a specific inability to see other minds. In a similar sense, Wimmer and Perner's test with ordinary three- and four-year-old children, seems to indicate that ordinary non-autistic three-year-old children are unable to see other people as minded.

However, the false belief task has received some criticism. In their article "Two Reasons to Abandon the False Belief Task as a Test of Theory of Mind" (2000) the psychologists Paul Bloom and Tim P. German argue that

the results of the false belief task cannot be taken for granted. They suggest that there is a problem in the way the false belief task is supposed to point at a general mental mechanism of mindreading or the lack of such a mechanism. This problem is connected with the fact that researchers often conclude that children under three as well as persons with autism fail the false belief task and that they therefore lack a theory-of-mind function. According to Bloom and German, such a comparison of the results are problematic because it suggests that a normal three year-old who fails the task is *similar* to an autistic child who fails the task.

Normal 3-year-olds and older children with autism both fail the false belief task, but, in all interesting regards, *normal 3-year-olds are nothing like older children with autism* [...]. Normal 3-year-olds are far superior with regard to communicative and linguistic skills, the ability to pretend and understand the pretence of others, and the ability to engage in, understand and manipulate the actions of others. This is a severe problem for any theory that lumps the two groups together as individuals who lack theory of mind. (Bloom and German 2000, p. B29)

Bloom and German argue that it is problematic to compare a normal three-year-old with an autistic child simply by measuring their failure or success in the false belief task. Such a comparison creates the impression that an ordinary three-year old child who fails the false belief task and an autistic child who fails the task have *similar* problems.

However, as Bloom and German note, a normal three-year old child is highly social even if the child might not pass the false belief task, while a child with autism has severe social problems. The question is then how the results should be explained.

One reason why both ordinary three-year-old children and children with autism fail the false belief task may have to do with it being constructed in a rigid and intellectualistic manner. Bloom and German point out that the false belief task, despite the appearance of simplicity, requires an advanced capacity to concentrate and remember story lines.

To solve it [the false belief task], the child has to follow the actions of two characters in a narrative, has to appreciate that Sally could not have observed the switching of the chocolate, has to remember both where the chocolate used to be and where it is at the time of the test, and has to appreciate the precise meaning of questions (for instance that it means where will Sally look, not where she *should* look). (Bloom and German 2000, p. B27)

As Bloom and German argue, the reason why ordinary three-year-old children fail the false belief task may be that it requires quite high attentional and linguistic skills. That young children fail the task does not necessarily point towards a theory-of-mind deficit. It might be difficult for the children to remember long story lines, it might be difficult for them to remember where a thing was placed

and it might also be difficult for them to understand certain words.

Another problem with the task is that it appears to be based on the idea that we understand other people's beliefs by *observing* and *reflecting on* their behaviour. But it is not usually in the form of such passive observation and reflection that we come to understand other people's mistakes or thoughts or intentions. Even if we occasionally do so this is by no means the major way in which we understand mistakes. On the contrary it is usually because we are *involved* in doing something *with* others that we also realise that someone makes a mistake. This indicates another reason why the children under three fail the false belief tasks even though they might have no problem in understanding mistakes in real life. They are placed in a situation where they should describe a situation rather than themselves take part in the situation. It is not self-evident that it is as easy for a three year old to *describe* a situation as it is to *respond* in a situation she is involved in herself. If a child is not used to commenting and describing a situation this can be a difficult task.

However, Bloom and German note that not all perspective-taking tasks are as rigid in their character as the classical false belief task. There have also been modified perspective-taking tasks where younger children succeed in the task. One such task was built up by the psychologist Daniela K. O'Neill (1996). O'Neill built up a task where the children are more engaged in doing things themselves rather than merely having to watch a scenario. The children are also involved with their parents rather

than with a researcher (i.e. a stranger). Bloom and German describe the task as follows.

[...] in an elegant study by O'Neill (1996), 2-year-olds observed as an attractive toy was put on a high shelf. As this happened, the child's parent was either present or absent. When later asking for help in retrieving the toy, the children were more likely to name the toy and gesture to the location when their parent had not been present to witness the placement of the toy than if their parent had been present. This suggests that they modify their behavior according to the knowledge states of other people (i.e. whether or not their parent possesses a given belief), and that they have a tacit appreciation of the circumstances under which beliefs are formed. (Bloom and German 2000, p. B29)

One can see here that when the test is modified so that it largely looks like an ordinary situation where the child is actively involved in doing something fun with family members, children of a much younger age act in a manner that indicate that they can take into account another person's knowledge or lack of knowledge of the situation.

However, the psychologists Vasudevi Reddy and Paul Morris (2009) point out that structural criticism of the false belief task can actually work as a *defence* of the idea that interpersonal understanding consists in a theory-of-mind function. In this sense O'Neill's more engaged false belief task, where the two-year olds ask for help to get a toy from a shelf, can be taken as support for theory-of-mind theory. Reddy and Morris write:

This early experimental test-passing could be interpreted as validating and being validated by the naturalistic data, and providing a new watershed for the development of false-belief understanding. In other words, the naturalistic data and the new experiments could be seen as coming together to simply lower the critical age for false-belief understanding to just after 12 months, and as providing new and dramatic evidence for the reality of a (now non-verbal) 'Theory of Mind'. (Reddy and Morris 2009, p. 95)

Reddy and Morris' point is important. It is not clear that structural criticism of the false belief task means that the theory-of-mind theory is flawed. Criticism of flaws in an empirical test does not necessarily always disqualify the theory that the test is supposed to prove. Both O'Neill's criticism and Bloom and German's criticism of the classical false belief task are specifically structural criticisms of the task itself. They are not explicitly questioning the idea that there is some such thing as a theory-of-mind function, but they are questioning the highly intellectualistic character of the classical test as well as the assumptions that are drawn from the results. Bloom and German also claim that the results of O'Neill's test indicate that normal children do have a theory of mind while children with autism do not have a theory of mind. They write:

A more promising analysis is that some individuals with autism fail the false belief task because they lack the capacity to acquire a theory of mind. In contrast 3-year-olds might fail the false belief task because of

general task demands, because they don't have a grasp of false belief, or both. But they surely have a 'theory of mind', in the general sense of having sophisticated ability to reason about the mental states; this is precisely why they differ from autistic individuals in the social, communicative, and imaginative domains. (Bloom and German 2000, p. B29)

Even though Bloom and German are critical of the false belief task, they are not questioning the idea that interpersonal understanding rests on a theory-of-mind function. They also claim that persons with autism lack a theory-of-mind function.

However, the problem with the classical false belief task is not only that the task is constructed in a too intellectualistic fashion, but the task reflects certain conceptual assumptions that underlie theory-of-mind theory. One such conceptual assumption is the notion that our understanding of other people in an essential sense is based on a third-person perspective where we observe other people instead of being involved in interaction with others.

If we now turn to reflect on children with autism, there seem to be other reasons why they fail the task. One aspect that may affect an autistic child's capacity to succeed in the task is that these children often have many kinds of attentional problems. They may suffer from various kinds of syndromes such as over- or under-sensitivity to sounds, light, touch, smells etc. Children with autism can also tend to get, so to say, stuck in their attention. A child with autism might seem to have a normal comprehension of

words, but the child might have a tendency to get stuck with certain words when listening to story lines. The child might also get stuck when looking at certain objects, the child may take an interest in things or expressions that are not perhaps the central aspect of the story. Further, often children with autism do not like to play with dolls. If that is the case it is not a very promising start to be placed in front of dolls in a test situation. From the descriptions of the classical false belief task it appears that no such individual syndromes or personal character traits are taken into account. My point here is not to say that actually children with autism *do* understand other people's belief. On the contrary, it is clear that it is often difficult for children with autism to understand others. I am merely claiming that since the false belief task is constructed in a manner that does not take into account the autistic child's individual syndromes, it becomes unclear how one should interpret the results of the false belief task.

Another thing I have tried to argue so far is that there may be quite different reasons why an ordinary non-autistic child under the age of three fails the false belief task and why a child with autism fails the task. This by itself indicates that the results are not as evident as they may seem. Still, even if the results of the false belief task are obscure, it is clear that it is often difficult for children with autism to consider other people's perspective. Park describes many instances where her daughter behaves as if she was unable to do so. Park here uses her daughter's real name, Jessy instead of "Elly" as in her first book.

Thinking of others, of course, is hard when you don't have a 'theory of mind' to allow you to see something from another point of view. Even in the unemotional, physical world, Jessy can't do this. She locks the door behind her when she leaves for work, even though she knows I'm still inside and there's no need to. She scrapes the ice off the windshield on the passenger's side, her side, leaving the driver's side obscured. She thinks I can see what she sees; if she knows something, she thinks the person she's talking to knows it too. (Park 2001, p. 148)

These patterns of behaviour suggest again that Jessy is unaware of other people's perspective. However, can one from cases like these draw the conclusion that children as well as adults with autism lack a theory-of-mind function?

Why do ordinary children learn to scrape the whole window and not merely the passenger's side? I am not convinced that they learn it by reflecting on other people's minds. For ordinary children it might not even occur to them that they could scrape *only* the passenger's side. Of course ordinary children can behave in selfish ways, as anyone can, but it is often as natural for them to do things in a way that takes others into account. My suggestion is that much of the things children learn to do include other people in a self-evident way that does not depend on a need to reflect on other people's minds. Sometimes questions about another person's perspective do come up, but there is no general basic question of "learning to see other minds" that non-autistic children apply at every instance when they take others into account. It is then also

problematic to maintain that there is some such general mindreading ability that a child with autism lacks when the child only scrapes the passenger's side of the window.

The false belief task is based on the assumption that other people in an essential sense are mental beings whom we need to learn to interpret. I have argued that children do not grow up to see that there is some such unitary thing as "intentions" or "beliefs" or "perspectives"; they grow up to comprehend and respond in various sorts of meaningful situations when being involved with other people. Often there is no question at all of other people having separate perspectives. Often we simply do things together because that is how things are done. Often we also spontaneously take each other into account. Sometimes, however, we do think about another person's perspective, but this is no general matter.

It is an important part of growing up that we learn to acknowledge others and to care for others, and it is important to learn that human life is a life full of mistakes and failures. But while theory-of-mind proponents think of this as a cognitive matter of learning to read other minds, I argue that we learn to respond to each other in a social and moral sense. This is largely something we grow into spontaneously as we grow up with others. Learning to take other people into account, learning to see that people (both others and myself) are not perfect, learning to help others, learning to see that they can be mistaken as well as that they can behave stupidly and arrogantly etc. is integral to how we respond to each other and integral to how we talk. It is integral to the fact that we help each other, that we

show each other where to look for something, it is integral to the fact that we accuse each other for things, that we get angry in disproportionate ways, that we sometimes later on apologize for our own behaviour, that we laugh at our own behaviour as well as tease others for their mistakes and so on. And it is also entwined with learning that sometimes it is irrelevant who is right or who is wrong, sometimes you just have to stop quarrelling. Such situations are very common for children. It is a very common thing for a child to lose things or to realize she was wrong about something. And it is quite common that parents get angry at their child for messing up and losing stuff etc. It is also common that parents help their children and that children help parents. It is an equally everyday fact that parents can't find things and rumble around searching for socks or gloves or boots.

The false belief task creates the impression that children learn to apply a general method of reflection when they learn to acknowledge others in their actions, as well as the impression that it is such a general capacity for "mindreading" that a person with autism lacks. I have argued that this impression is created by the intellectualistic character of the task.

3.5 Theory of mind and autistic children's language problems

A further aspect of autism that seems to point towards an inability to see that other people have minds is the fact that children with autism often have severe language problems.

This has also been taken to support the theory-of-mind theory. Referring to Leo Kanner's studies, Hobson gives an overview of what kinds of peculiarities can show in the way a child with autism speaks.

Leo Kanner (1943) noted that besides a lack of communicative speech that may amount to muteness, the autistic child commonly displays echolalia (an 'echoing' of words or phrases the child has heard spoken by others, either in the immediate or more distant past), confusions in the use of personal pronouns 'I' and 'you', idiosyncratic utterances that can only be understood with reference to the contexts in which the child acquired the words, and a literalness of speech that seems to show a restricted grasp of connotative meanings. Amongst other features one may add to this list are abnormalities in the tone and rhythm of speech (which may be flat and monotonous, or sing-song) and difficulties in initiating or sustaining conversation with someone else, partly through insensitivity to the knowledge and interests of the listener [...] (Hobson 1993, p. 165)

Hobson suggests that these features are reflective of the child's great social difficulties.

Rather than underlying the handicaps of autism, language might constitute an especially refined and vivid reflection of the children's limitations in interpersonal relatedness and understanding, and more specifically in their notions of sharing, referring, and communicating. (Hobson 1993, p. 166)

One could easily take the tendency for echolalia as well as the tendency to mix up the use of the words “I” and “you” to indicate that the child with autism is unable to see that other people have different perspectives and that the child is actually not communicating with others even though he or she is talking. However, Hobson points out that this is jumping to conclusions. He writes:

The first lesson to be learnt from this research is that we should be wary of supposing that autistic children are non-communicative. What might at first appear to be failures to use speech and gesture for communicative purposes may often turn out to have significance for the children’s interpersonal transactions. For example, the autistic child’s use of idiosyncratic or ‘metaphorical’ expressions may reflect an attempt to communicate, but one that is ineffective in so far as the message is not adapted to the listener’s perspective⁴⁰ Echolalia may seem (and sometimes be) meaningless, but it can also serve a variety of purposes such as to curtail a social exchange⁴¹, to maintain an interaction in the face of a failure to comprehend another person⁴² or to fulfil a variety of other functions such as requesting, protesting, affirming and so on⁴³. Incessant and repetitive questioning may be intended to initiate or to maintain social contact rather than to

⁴⁰ Hobson refers here to Wetherby (1986).

⁴¹ Hobson refers here to Shapiro (1977).

⁴² Hobson refers here to Fay (1973).

⁴³ Hobson refers here to Prizant & Duchan (1981), and Prizant & Rydell (1984).

request information ⁴⁴ . Correspondingly, autistic children vary in their wish or ability to communicate with different individuals, and may be more communicative with teachers than with children or with familiar than with unfamiliar adults⁴⁵. (Hobson 1993, p. 174)

According to Hobson, one cannot draw the conclusion that an autistic person is unaware of other people, and does not communicate with speech, simply because the person has a tendency to speak in terms of echolalia or is unable to distinguish between the words “I” and “you” or because the person’s speech in other ways *appears* to be non-communicative. The education and disability researcher Douglas Biklen (2005) also notes that echolalia sometimes can have to do with the tendency of some people with autism to get stuck with certain words and expressions. He quotes an autistic person named Sue Rubin who describes such situations:

She describes herself as having ‘obsessive or compulsive behavior’ where she gets ‘stuck with certain thoughts and actions’. With echolalia, she explains, ‘I say a word or sound and am unable to switch it off or change to a different sound’. Yet, when

⁴⁴ Hobson refers here to Hurtig, Ensrud, & Tomblin (1982).

⁴⁵ Hobson refers here to McHale, Simeonsson, Marcus & Olley (1980) and Bernard-Opitz (1982).

attending a class where she is 'cognitively engaged,' her echolalia disappears. (Biklen 2005, p. 43) ⁴⁶

Rubin's description of how she sometimes can get stuck in echolalia suggests that the echolalia need not have to do with an inability to understand other people's words. Rather it can be a kind of obsessive behaviour that the person gets caught in but that the person also, sometimes, can manage to overcome. However, such a tendency to get stuck on words can probably also affect a child's ability to learn to speak.

Hobson's description above also suggests that there are many aspects to what it means to talk with another person. Sometimes we try to initiate conversations as a way of being together. Sometimes a failure to understand can show in a person clinging to certain words or repeating the same question several times, in order not to have to show the failure to understand. Sometimes also disappointment can show in repeated questions. And, importantly, our ways of talking, our wish to initiate conversations, our wish to ask or explain etc. often differ considerably depending on *whom* we talk with, whether it is a familiar person or not. Hobson's descriptions suggest that one should be careful about drawing too definite conclusions about what a person with autism can or can't understand about others. In many situations it may be difficult for the person with autism to see another's perspective, but this is

⁴⁶ Biklen refers here to Rubin, Biklen, Kasa-Hendrickson, Kluth, Cardinal & Broderick (2001).

not necessarily some *all-encompassing* feature meaning that the person has no sense for other people whatsoever. Nor does it prove that the person lives in a solipsistic world isolated from others. An autistic child's inability to comprehend how the meaning of situations vary depending on whom one talks to might be reflected in a confused use of personal pronouns, it might also be a feature that makes it difficult for the child with autism to understand and respond to more complicated situations as when several people talk with each other, or when people talk about the future or about situations that are not present, but these problems of understanding need not be absolute.

In another book by Clara Park, *Exiting Nirvana, A Daughter's Life with Autism* (2001) Park reflects on her autistic daughter as an adult. Here one can again see traits that seem to fit in with the theory about autism being a theory-of-mind deficit. She describes, among other things, certain kinds of peculiarities in her daughter's development of language.

She speaks of Miranda, her brother's daughter, and someone asks, 'Who's Miranda?' She hesitates. Then, slowly and carefully, she replies: 'I...am...my niece.'

Be assured, Jessy knows she isn't her niece. Although eager psychoanalysts for years took pronominal reversal as evidence of 'early ego failure,' Jessy has anything but a weak ego. [...] Jessy has even more trouble with 'we' and 'our' and 'us,' with 'they' and 'their,' with 'his' and 'hers,' even with 'he' and 'she.' I hear her answering the phone, groping for the

words to tell a caller her father's not home. It's going slowly, so I try to help. 'Say he's not home and can you take a message.' Jessy alters the primary pronoun; she's learned that much. But what results is this: 'she's not home, and can you take a message.' (Park 2001, pp. 42-43)

From Park's description one can see that confusing personal pronouns does *not* necessarily mean that one does not know the difference between persons. Still, the tendency to confuse pronouns does point at some kind of difficulty to grasp the variable character of how we address persons, how we talk about others, how we talk of relationships etc. According to Park, Jessy tended to understand words referring to "objects" much more easily than words referring to persons.

A giraffe is a giraffe wherever you find it; a rectangle is a rectangle. Not so with nouns like 'teacher,' 'friend,' 'sister.' My teacher may be your sister, or her friend. It was not until late in Jessy's teens that we could teach her, with charts and written examples, the simple words for generational relationships. (Park 2001, p. 47)

In *The Siege* (1967) Park also writes:

[...] there were no limitations on the number of nouns she could acquire. But even among nouns, the easiest words to learn, there were limitations on kind. She could learn immediately a word like 'igloo' and remember it, although its relevance to her own experience was nonexistent. She could learn and

accurately apply the words 'oak,' 'elm,' and 'maple.' Yet words which were, one would think, much closer to her experience she could not understand or learn. Such terms as 'home,' 'sister,' 'grandmother,' 'teacher,' 'friend,' or 'stranger' were beyond her at five; 'friend' and 'stranger' are beyond her today. Proper names she acquired with a slowness that seemed clearly related to the weakness of affect. (Park 1967, pp. 200-201)

A bit further Park continues (here using the name Elly).

As we observed Elly's developing speech, it seemed divided into words she could learn instantly once they were pointed out to her, and words she could not learn at all. For a long time there seemed to be no middle ground. What she was able to grasp were absolute terms, whether concrete or abstract—those that reflected concepts that could be defined and understood in themselves. 'Box,' 'cat,' 'giraffe,' 'rectangle,' 'number,' 'letter'. What she could not understand were relational terms—those that must absorb their full meaning from the situations in which they occur—situations in which the human element plays a part. [...] 'Teacher' is a word which, like 'man', is the product of abstraction, but it is first learned in a relational situation: 'my teacher'. Similarly for 'sister', 'friend' or 'home'. It is characteristic of the average child that he learns concepts best in situations in which he can find a personal relation. With Elly, the personal relation seemed at best irrelevant, at worst a hindrance. (Park 1967, p. 204)

It is tempting to think that these difficulties to learn words that are connected with human relationships point at Elly's (that is, Jessy's) inability to be aware of other people's minds and that it reflects mindblindness. However, think of how a child normally learns to use words like "mom" or "dad". They learn them in connection with doing such things as constantly following their parent, requesting comfort, with asking or notifying the parent about something, they learn it as part of expressing their delight as a greeting when "mom" or "dad" comes home etc. Children also quickly learn the name of their siblings in connection with quarreling or teasing or playing with each other. For instance expressions like "Mom! Alexandra took my lego car!" is a very common way of using names. Children do *not* ordinarily learn such words because they learn to read minds, they learn these words because they are in constant interaction with others. If a child for a long time while growing up is incapable of such spontaneous ways of being with and responding to its family members, it will also affect the child's ability to use words like "mom", "dad" or names of siblings. Again, that Jessy did not use these personal words does not point to a theory-of-mind deficit, it points to how severely unresponsive she was as a young child.

However, Park also describes another pattern in Jessy's behaviour that might explain why it was difficult for Jessy to understand personal pronouns. She describes Jessy as if it was difficult for her to shift her attention from one thing to another. This showed when Jessy was a baby, in the sense that she could keep doing one simple thing for hours,

like snaking a chain in her hand. The same rigidity shows when Jessy is older among other things in that she finds it easier to understand words that refer to objects than words referring to persons. But it is also reflected in the fact that Jessy likes order; she does not like it when plans are changed suddenly. Jessy also likes catchy phrases, phrases that are often repeated, standing phrases:

When she watches TV, what is she watching for? Not the content, whatever that may be; not even the pictures. What she's listening for, what she *hears*, is what we've named 'intransition phrases': 'Coming up next,' 'Don't touch that dial,' 'Hold everything,' 'Stay tuned,' 'Be right back.' She is delighted to write them down for me. She has identified twenty-six. (Park 2001, pp. 50-51)

There seems to be a multitude of patterns in how Jessy talks and how she takes an interest in things. She does seem to have certain kind of problems with taking other people's perspective. But it is important not to define certain confused patterns of speech or surprising ways of behaving as if they necessarily mean that the autistic person completely lacks a comprehension of other people or of the situation in question. Baron-Cohen's way of talking about autistic persons as *mindblind* suggests such a complete and all-encompassing inability to understand other people. Further, I also think it is good to see that a person with autism does not only have problems and deficiencies but can also come to enjoy doing things and

can come to enjoy life in new and sometimes surprising ways.

3.6 Theory of mind and autistic persons' difficulties to understand irony

In this last part of the chapter I shall discuss one more problem of understanding that, according to proponents of theory of mind, is connected with autism, namely the inability to understand ironic speech.

According to the cognitive neuroscientist Francesca Happé (1993), people with autism often have problems understanding metaphor and irony. This, according to her, has to do with the same mindreading deficit that is reflected in the autistic child's inability to manage the false belief task. Happé has shown there to be a correlation between how well or badly certain autistic people accomplish false belief tasks and their ability to understand metaphor and irony. From her research one gets the impression that there is a connection between a capacity to see that other people have inner, invisible intentions (which, according to her, shows in the capacity to pass the false belief tasks), and a capacity to understand metaphor and irony.

Research into the autistic child's theory of mind⁴⁷ has found a severe impairment in most autistic subject's ability to comprehend another person's false belief. [...]

⁴⁷ Happé refers here to Baron-Cohen, Leslie & Frith (1985).

This impairment should have serious consequences for communication, if, as Sperber and Wilson (1987) claim, 'communication exploits the well-known ability of humans to attribute intentions to each other.' (Happé, 1993, pp. 101-102)

Happé continues:

If most autistic individuals cannot represent a speaker's intention, then communication should break down most noticeably where the speaker's attitude must be taken into account in modifying the literal meaning of the utterance. [...] It is widely reported⁴⁸ that even the most verbally able autistic people fail to understand non-literal speech such as irony, joking and metaphorical expressions. (Happé 1993, p. 103)

Autistic speakers often use language in a rigid manner. What is especially difficult for autistic speakers, according to Happé, is the *intentionality* of language and thus how the meaning of our words may differ depending on how or why we say things. I agree with Happé that people with autism can have a rigid or unusual use of language and that it can be difficult for them to understand metaphor and irony. But the question is if these difficulties can be explained as depending on a mindreading deficit. The problem here lies in the idea that there exists a certain general feature of our language, its "intentionality". This aspect of language is what is thought to be the thing that autistic persons cannot grasp. According to Happé it is an

⁴⁸ Happé refers here to Happé (1991) and Tantam (1991).

important feature of language that we can mean different things by what we say even though we use the same expression. She says that sometimes we mean things “literally” but sometimes “the speaker’s attitude must be taken into account” in order for us to understand what the speaker means. According to her, this shows that we must be able to see others as minded, as having intentions, in order to understand how the meaning of words can vary. Even though I agree with Happé that we can sometimes distinguish between “literal” and “non-literal” speech, I do not think this is a fundamental distinction of how language has meaning. Neither am I convinced that such a distinction helps us understand how or why the meanings of our words differ. Nor do I think the distinction helps us understand how humour and irony have meaning in conversations, or why a person with autism may find it difficult to understand humour and irony. Let me now examine my doubts.

Happé assumes that understanding irony and humour requires a special effort of mindreading. This creates, I think, a problematically intellectualistic impression of humour and irony. Another way to think of the meaning of our words is to look at how our ways of talking are an integral part of our daily lives with each other. Children grow up and learn to acknowledge others by constantly doing things together with others. In this constant interaction they also learn to respond to each other; help, quarrel, criticize, protest, apologize and so on. This also means that there is no fundamental step of realizing that words can mean different things depending on the

person's "inner intentions". Words mean different things depending on what we are doing and why we are talking with each other and how we are related to each other. And our sensing the difference in the meaning of what people say is so integral to how we sense a meaning in the specific situation that there usually is no question of the person's words perhaps meaning something else than we take them to mean. It is also as a part of such daily engagement that we learn to play with words, expressions, the tone of voice, manners etc. However, if a normal responsiveness and spontaneous engagement in various situations is broadly lacking in a child's way of being, it will also affect the child's sensitivity to what people say as it will affect the child's ability to comprehend situations that it is not used to.

Instead of taking an autistic speaker's difficulty understanding irony as pointing towards an inability to read minds, I think there is no single answer to this problem. However, one answer perhaps has to do with the place humour and irony have in our ways of talking with each other. As Hobson notes, children with autism seldom chatter spontaneously, seldom initiate conversations and are not good at sustaining a conversation. (Hobson 1993, p. 168) Irony and joking are centrally things that take place in such kinds of spontaneously elicited conversations and chats; it is part of the ways we enjoy conversations, how we continue them and go along in them. If it is difficult for a child to comprehend such kinds of spontaneous conversations, it might also be difficult for the child to understand irony. The inability to understand irony can

then be seen as a reflection of the person's large and long lasting problems with growing into a normal kind of responsiveness to others which includes the person's ways of becoming engaged in spontaneous conversations.

Hobson mentions a study by Ricks and Wing (1975) that illustrates how a child with autism may answer questions:

One was the following question-answer sequence:
 Question: 'What did you have for dinner?' Answer:
 'Meat and cabbage and potatoes and gravy and salt
 and jam tart and custard and orange juice and cup of
 tea.' (Hobson 1993, p. 170)

The child here gives a precise list of absolutely everything that he has eaten. His answer reflects his incapacity to sense the contextual meaning of the question. Usually when we ask another person what he has had for dinner we do not ask for a complete list of what has been consumed. Our ordinary ways of answering and asking things are entwined in a meaningful social context. In order to be able to understand what is a relevant answer and what is irrelevant, we must be accustomed to various ways of talking about dinners and eating; ways of talking where certain things *are* the "dinner" and certain things are simply side issues connected with the dinner and not relevant to mention. The child's way of giving a long detailed list of everything he ate, including custard and orange juice and tea, can be seen as reflecting his larger patterns of difficulty in engaging with others, such as for instance being involved in daily conversations about dinner. Such kinds of long lasting and broad social

difficulties can affect a child's ability to spontaneously understand questions in various contexts and to see what is a relevant answer in a context. It can affect the child's ability to flow along in discussions, to react to certain tones, to see certain relationships between people, and it can surely affect a child's capacity to understand irony.

From Happé's argument that autistic people do not understand irony it also appears as if we, "ordinary people", always understand irony. But people's capacity to understand irony is much more variable than Happé claims. This has to do with understanding irony requiring a deep acquaintance with various ways of living and talking, as well as often also an acquaintance with the person who is talking. Understanding or not understanding irony and joking is often connected with closeness to others. In some families joking is an almost constant part of their manner of conversing. It can also be a common pattern in some families that they become ironic towards each other when they quarrel. In other families a calmer and more serious way of talking forms the days. As Gilbert Ryle writes: "For one person to see the jokes that another makes, the one thing he must have is a sense of humour and even that special brand of sense of humour of which those jokes are exercises." (Ryle 1949, p. 54) Understanding irony also has to do with the fact that it takes time to learn to see certain patterns in your close ones' speech and manners. Becoming skilled at irony can be entwined with getting used to teasing your brother or your sister or the kids at school. Slowly many children get more and more skilled at this. Our use of irony as well as joking

also depends on how we feel at home in talking with others. It is difficult to try to be fun when you feel at unease with the people you talk with. In such situations all your words can feel clumsy and out of place. Irony also demands a high skill in using words. It is for this reason very difficult both to joke and be ironic in a language you do not speak as your mother tongue. The point in trying to describe these various aspects of irony is, that it is problematic to portray irony as if it could be understood by a picture of language as consisting of two mental beings communicating certain inner intentions. This is as problematic as portraying "false beliefs" as something unitary and inner in the mind.

Happé's claim that autistic persons do not understand irony because they have a mindreading deficit is not only based on a philosophical misconception of what it means to understand other people, but her reflections are also too absolute. A skill that even for many ordinary grownups is difficult to manage, when lacking in an autistic person, is taken as defining the autistic person as mindblind. I think it would be better again to see that humour is not an all or nothing affair. A person with autism is not necessarily without humour in some absolute sense any more than an ordinary person *has* humour in some absolute sense, but it can be the case that the autistic person does not manage to understand subtle forms of humour, or forms of humour that require a high sensitivity for the use of words. In the case of Jessy (i.e. Elly) described above, she loved to be tickled. This is already, I would say, an expression of a sense of playfulness and humour. And it is a kind of

playfulness that gets its meaning out of a sense of bodily closeness, not an intellectual form of humour. This sense for affection Jessy and her mother eventually developed into peek-a-boo games, and eventually Jessy also came to joke independently and more actively with her family. Jessy responded more and more to her parents and siblings in various ways; by giving a hug, by enjoying certain mutual games, and by joking with her family in turning the light off at supper and occasionally by getting angry. It was not because she learned to read minds that she began to joke; it was as part of a pattern of developing responsiveness and ability to communicate and thus more advanced enjoyment in being together with her close ones that the joking also became a part of her ways of being with her family.

However, even if I have argued that Happé's idea that the inability to understand irony is due to a theory-of-mind deficit is problematic, this does not mean that people with autism have no problems understanding others or being involved in conversations. I am merely claiming that such problems cannot be understood from the perspective of theory of mind. And I am also maintaining that the empirical research that is based on the theory about autism being a theory-of-mind deficit is one-sided in its focus. This has to do with the whole perspective of theory of mind being built on a misconception of what it means to understand another person.

3.7 Conclusion

My intention in this chapter has been twofold. One aim has been to discuss and probe the idea that persons with autism lack a theory-of-mind function. However, another aim, which connects with the other chapters in this thesis, has been to once again question the relation between empirical research and theory-of-mind theory.

There are many traits in autism that may appear to support the theory that autism is due to a theory-of-mind deficit. Children with autism can be largely unaware of other people, they can have extensive language problems, they are often more or less unable to play with others, they are not good at pretend play etc. Such problems can seem to point towards an inability to see that other people have minds. I have, however, questioned the theory that autism consists in "mindblindness". Among other things, I have argued that the image of autism as mindblindness is created by a one-sided and restricted focus when discussing the responses of children with autism. I have also argued that theory-of-mind theory is based on an intellectualistic conception of normal interpersonal understanding.

Chapter 4: The relationship between empathic imagination and compassion⁴⁹

4.1 Introduction

In philosophical and psychological contexts empathy is often talked about as a general capacity to imagine what another person feels or thinks or as a capacity to imagine oneself in the other person's situation. This capacity is sometimes described as "putting oneself in the other's shoes". It is suggested that the capacity to imagine what the other person feels is a fundamentally important feature that enables us to understand other people. Often this imaginative capacity is considered to be a phenomenon that can concern any kind of emotional state, regardless of whether the other person is happy or sad or angry or afraid etc. According to several theorists, there is an especially important link between the imaginative-emotional function of empathy and our willingness to help other people. It is claimed that empathic imagination makes us emotionally moved by another person's situation, and this is then thought to motivate us to help the other.

In this sense the concept of empathy, in philosophical and psychological theories, often has a cognitive meaning that is distinct from compassion, even though it is suggested that there is an important link between empathy and compassion. The idea that there is an empirically observable link between empathic imagination and

⁴⁹ Parts of this chapter have been published in Gustafsson (2009).

compassion also appears to be reflected in our everyday experiences. We are often emotionally moved by the sight of another person's suffering. We can find it unbearable to look at an injured person. It is as if *we* feel the other person's suffering. And often we also reflect on other people's suffering, we worry, etc. Imagination and emotional reactions are in this sense often a central part of compassion. Often it is also the case that if we have experienced something similar as another person we also feel we understand how the other person feels. A common way of talking about compassion is to say that we feel *with* the other person. Sometimes we might also say that we *identify* with the other. These ways of responding emotionally to other people's suffering seem to support the suggestion, originally made by David Hume and Adam Smith, that compassion is based on an analogical method of imagination, a form of explanation that is often invoked in theory-of-mind theories. In this chapter I will discuss the relation between empathic imagination and compassion.

I begin the chapter by discussing experimental research on empathy that is considered to prove that there is a causal link between empathic imagination and compassion. I claim that there are certain problems connected with the experimental research. Since the experimental context is highly reduced it becomes unclear in what sense the participants can be said really to understand that another person is suffering. It is central for how we understand another person's suffering that we can see that the suffering affects the person's life in some sense. In some of the experimental situations, however, the painful

experiences have no consequences; they are not related to injury or illness. I also claim that the instructions that the experimenters give to the test subjects are leading. Thereby it also becomes unclear how the responses should be understood. I further argue that the fact that the researchers construct such restricted research settings is reflective of the assumption that compassion depends on a cognitive capacity to use our imagination in an analogical sense.

After having discussed experimental research, I discuss several biographical descriptions of suffering as well as of compassion. These biographical descriptions reveal how experiences of suffering affect interpersonal relationships in many ways. The descriptions also reveal how our difficulties in understanding another person's suffering are of a quite different character than is assumed in the psychological and philosophical theories on empathy.

In the last part of the chapter I discuss the idea that empathic imagination can be used for both good and evil purposes. This idea rests on a conception of interpersonal understanding as a neutral cognitive faculty. I argue that such a conception of interpersonal understanding is problematic. There is no general neutral way of understanding other people's suffering. The ways we think about suffering, the ways we know things, reflects our attitude towards the suffering person. In this sense knowledge and imagination cannot be separated from the fact that we are responsible for each other's life.

4.2 The relation between empathy and compassion

The psychologist Ezra Stotland (1969) describes empathy in the following way:

A mother will share the joys and sorrows of her children; friends often feel each other's emotions; the sight of a sick or injured person will sometimes upset us; and we are sometimes elated at another person's success. Our sharing of the feelings of another does not, however, *necessarily* imply that we will act or even feel impelled to act in a supportive or sympathetic way when we are reacting to another's sorrows. [...] Nevertheless, on other occasions, a person may be moved by another's pain to help the other, or to help another attain and sustain a happy experience. In short, sharing another's feelings should be distinguished from acting sympathetically and helpfully towards him. [...] The phenomenon referred to [...] can be described as 'empathy,' [...] It is an observer's reacting emotionally because he perceives that another is experiencing or is about to experience an emotion. (Stotland 1969, p. 272)

Indeed there is much that seems correct in Stotland's description above. We do share the joys and sorrows of our loved ones, people do often get upset by the sight of an injured person etc. We are often emotionally moved by other people. It seems also correct to say that such emotional movement is not always expressive of compassion. According to the social psychologist C. Daniel

Batson (1991), there is a connection between empathy and altruism (or in other words compassion). Batson maintains that empathy is linked with altruistic motivation to help the other person.

Empathy is an other-oriented vicarious emotion produced by taking the perspective of a person perceived to be in need. It is distinct from personal distress. The magnitude of empathic emotion is a function of the magnitude of the perceived need and the strength of the perceiver's attachment to the person in need. [...] Empathic emotion evokes altruistic motivation to have the other's need reduced. [...] the goal of this motivation is to increase the other's welfare, not one's own. (Batson 1991, pp. 89-90)

Batson further claims that it is of central importance for empathy that we are capable of adopting the other person's perspective:

[...] perception of the other as in need is a necessary but not sufficient condition for the experience of empathy. Also necessary is adoption of the other's perspective, perceiving the situation from the point of view of the other's wants and desires and imagining how the other feels about the situation.

Adoption of the other's perspective requires considerable cognitive sophistication. It requires the ability to view the need situation from a point of view other than one's own and the ability to imagine how someone else is feeling or will feel. These inferences must often be made using limited information about

the situation and limited verbal and nonverbal cues.
(Batson 1991, p. 226)

Martha Nussbaum (2001) also defines empathy as an act of imagination that is related to compassion.

[...] compassion is distinct from empathy, which involves an imaginative reconstruction of the experience of the other. [...] empathy is like the mental preparation of a skilled (Method) actor: it involves a participatory enactment of the situation of the sufferer, but is always combined with the awareness that one is not oneself the sufferer. This awareness of one's separate life is quite important if empathy is to be closely related to compassion: for if it is to be for *another*, and not for oneself, that one feels compassion, one must be aware both of the bad lot of the sufferer and of the fact that it is, right now, not one's own. (Nussbaum 2001, p. 327)

A few pages later Nussbaum continues:

[Empathy] is a very important tool in the service of getting a sense of what is going on with the other person, and also establishing concern and connection. [...] By reconstructing in my own mind the experience of another, I get a sense of what it means for her to suffer that way, and this may make me more likely to see her prospects as similar to my own, and of concern in part for that reason. (Nussbaum 2001, pp. 330-331)

However, according to Nussbaum, there is no necessary connection between empathy and compassion. According

to her, empathy is a morally neutral mental ability that, even though it can be useful for compassion, can also be used for sadistic ends.

In short, empathy is a mental ability highly relevant to compassion, although it is itself both fallible and morally neutral.

Does empathy contribute anything of ethical importance entirely on its own (when it does not lead to compassion)? I have suggested that it does not: a torturer can use it for hostile and sadistic ends. (Nussbaum 2001, p. 333)

Stotland's, Batson's and Nussbaum's thoughts reflect several important aspects of compassion. As Stotland notes, we are often emotionally moved by another person's suffering. And, as Batson and Nussbaum point out, it is often the case that another person's suffering affects our imagination. However, all three also seem to share certain theoretical assumptions concerning empathy and compassion. One such assumption is that there is a kind of general cognitive capacity to imagine and to become emotionally moved by other people, that we can define as empathy. Another assumption is that there is a causal psychological link between such empathic imagination and our inclination to help others (i.e. compassion).

Both Stotland and Batson have also conducted psychological experimental research on the relation between empathy and compassion. The experimental results seem to indicate that there is a causal connection

between empathic imagination and our willingness to help other people.

4.3 Stotland's experimental research on empathy

The suggestion that there is a causal psychological link between empathy and compassion has been studied empirically in several psychological experiments. From the results of these experiments it seems that the existence of an empathic function is a well confirmed fact. According to Stotland, there is a causal relation between altruism (i.e. compassion) and our capacity to feel empathy. In order to test this hypothesis, Stotland and Stanley Sherman designed a psychological experiment. Stotland describes the experimental setup in the following way:

In general, the design of the study [...] involved inducing three different mental sets in the subjects as they observed another person undergo a painful, neutral, or 'pleasurable' experience [...] The first set was that as they observed the other, they were to imagine how they themselves would feel if they were in the other's position ('imagine-self' condition.). The second set was that they were to imagine how the other person felt ('imagine-him' condition). The remaining subjects were asked to watch the other person's physical movements very closely ('watch-him' condition). The general hypotheses were, first, that more empathy would occur in the imagine-him condition than in the watch-him condition; second, that more empathy would occur in the imagine-self condition than in the watch-him condition; third, that more empathy would

occur in the imagine-self condition than in the imagine-him condition. (Stotland 1969, p. 289)

The aim of Stotland and Sherman's experiment was to see how differences in perspective affect our capacity to feel empathy. The degree of empathic feeling was measured both physically and by answering questionnaires. The physical effect was measured by a vasoconstriction measure and by a palmar sweating measure. I will not now discuss all aspects of Stotland's experimental research but will focus on some details.

The result of these experiments was that the test-subjects in the two imagining conditions (imagine-self and imagine-other) reacted more strongly than the test-subjects in the watch-him conditions.

[...] on none of the physiological measures was there any significant differences among the watch-him—pain, watch-him—neutral, and watch-him—pleasure conditions, while there were differences on the physiological measures in the two imagining conditions. This indicates empathy is related to the set that the person has in viewing the other person. A 'superficial' set of just watching him does not lead to empathy. (Stotland 1969, p. 297)

From Stotland and Sherman's experiment one gets the impression that a certain form of imagination induces empathic feeling. Their experiment has also influenced further research on empathy. According to Batson (1991), the results of experiments carried out by Stotland and

Sherman support the claim that there is a link between empathy and compassion. Alvin I. Goldman (1995) also finds Stotland and Sherman's experiment convincing. According to Goldman, the experiment proves that the mechanism of empathic imagination can produce compassion. He concludes: "[Empathy] seems to be a prime mechanism that disposes us toward altruistic behavior" (Goldman 1995, p. 202). However, it is worth taking a closer look at Stotland and Sherman's experiment. Let us first look at the instructions that were given to the test-subjects in the watch-him condition.

Watch-Him Condition

In a few moments you will be watching the actual demonstration. While you are doing so, please watch exactly what the demonstrator does. You are to watch all of his body movements that you can see. Your job will be to watch his leg movements, arm movements, foot movements, head movements, hand movements. You are to notice anything that he does, whatever it is. (While you are watching him, don't try to imagine how you would feel in his place or how he is feeling. Don't think about how he feels or how you would feel. Just watch him closely.) (Stotland 1969, pp. 292-293)

There are some points about the experimental design that the experimenters seem to have neglected. It is plausible to suppose that the fact that the persons who watch carefully do not have strong emotional reactions reflects how they understand the request in the experimental situation. They are *asked* to put their feelings aside. That is, the questions

are leading. Similar problems with the experiment can be seen if one looks at the other two groups that were asked to use their imagination in the imagine-self condition or in the imagine-him condition. These were told the following:

Imagine-Self Condition

In a few moments you will be watching the actual demonstration. While you are doing so, please imagine how you yourself would feel if you were subjected to the diathermy treatment, whether it turns out to be painful, pleasant or neither. While you are watching him, picture to yourself just how you would feel. ([...] You are to react as if it were you who will have the experience that is pleasant, painful or neither.) While you are watching him, you are to concentrate on yourself in that experience. You are to concentrate on the way you would feel while receiving the treatment. Your job will be to think about what your reactions would be to the sensations you would receive in your hand. (Stotland 1969, p. 292)

Imagine-Him Condition

In a few moments you will be watching the actual demonstration. While you are doing so, please imagine how the demonstrator feels as he is subjected to the diathermy treatment, whether it turns out to be painful, pleasant, or neither. While you are watching him, picture to yourself just how he feels. (You are to keep clearly in mind that [...] It is he who will have the experience that is pleasant, painful, or neither. While you are watching him, forget yourself.) While you are watching him, you are to concentrate on him in that

experience. You are to concentrate on the way he feels while receiving the treatment. Your job will be to think about his reaction to the sensations he is receiving in his hand. In your mind's eye, you are to visualize how it feels to him to be the demonstrator in this experiment. (Stotland 1969, p. 292)

It is not evident that these two groups that were asked to use their imagination were merely (as Stotland claims) asked to use a certain cognitive technique that was directed at the self or at the person being in pain. That is, the experimenters thought they exhorted the test persons to use a neutral cognitive technique with two kinds of focus; one where the focus was directed at oneself, and the other where the focus was directed at the person in pain. According to Stotland, this difference in focus also reflects a difference in the emotional reactions. However, it is still possible that the test persons understood the questions in a different manner; as exhortations to take an egocentric or compassionate attitude. This means that the result according to which the test persons who were told to imagine, tended to react more strongly than the group that was told to watch, is not necessarily a result of their using a mental technique. Rather the different reactions might well be a result of how they understood the exhortations in the experiment in a certain sense, i.e. as an exhortation to feel self-centred nausea (or pleasure) or as an exhortation to care. The persons were asked to imagine in order to fulfil the demands in the experimental situation. This means that the exhortations become ambiguous. The test

subjects' strong emotional responses do not necessarily reveal how we might react in real life since the test persons are trying to fulfil the demands in the experiment. However, do we not sometimes also in real life tell people to use their imagination? Sometimes I might say to my child "Try to imagine how *you* would feel if you were in her situation!" Indeed we do say things like these, but they are in themselves moral exhortations, they are not neutral advice.

However, the problem with Stotland and Sherman's experiment is not only that the questions are leading. There are also certain theoretical assumptions concerning interpersonal understanding that seem to affect how the experiment is designed. These assumptions seem to be one reason why Stotland and Sherman do not notice that their questions are leading.

One assumption concerns the nature of the relation between seeing another person and imagining how another person feels. The assumption seems to be that when we see another person we see, so to say, merely the person's bodily surface. Therefore, in order to understand how another person feels when he is in pain we must use our imagination. The idea that we always have a deeper understanding of people when we imagine how they feel than when we look at them might seem natural against the background of a dualistic way of thinking about human beings. The researchers do not themselves say that they have a dualistic conception of human beings. Nevertheless one can see a tendency towards a dualism in their way of distinguishing between simply watching bodily

movements and, on the other hand, imagining another person's thoughts and feelings. The dualism gets constructed in the experimental design, in the very instructions given to the test persons. One can also see such a tendency towards dualism in the earlier quote where Stotland concludes that "A 'superficial' set of just watching him does not lead to empathy." According to Stotland, watching another person is more superficial than imagining how the other person feels. However, the character of watching that takes place in the watch-him condition is actually a very sophisticated form of seeing. It is not the normal, basic or the only way of seeing others. It is, as I have already argued, something the persons are *instructed* to do. In the "Watch-him condition" in the experiment the test-subjects are exhorted to adopt a fragmented way of looking at another person. The subjects are asked to look at leg movements, arm movements, etc. This fragmented way of looking at the movement of bodily parts in the "Watch-him condition" affects the test-subjects' capacity to perceive the other person's pain.

Ordinarily when we see that another person is in pain we see the human being's behaviour as a meaningful whole and as part of a meaningful context. This way of seeing other people is immediate. From such a perspective the most apparent way of describing a person who is in pain would be to say that he is grimacing in pain or that he is writhing in pain. Normally our seeing others is also inseparable from the ways they engage us. Think, for instance, about the way you look into the other person's eyes while laughing at a joke, or you look angrily at

someone, or you see that a person on a bike falls and you instantly go to help him, ask him if he hurt himself etc. None of these forms of looking or seeing have to do with observing bodily parts that are moving in certain ways; they are responses to the other person in a meaningful situation and as such also expressions of understanding.

However, there are also certain situations in real life where we can adopt a fragmented way of looking at others. For instance, if I have a bad knee and visit the doctor she might study the movement of my leg in a clinical way. In this sense there is a resemblance between certain medical contexts and the way in which the subject in the experiment is asked to observe bodily movements. However, a medical context is a special context, it is not a situation that reflects how we generally look at each other or how we ordinarily recognize that another person is in pain. The doctor's capacity to observe my leg movements is expressive of her skill as a doctor, and this skill is dependent on the fact that she already understands that my knee hurts. That is, the focused way of looking that a doctor can sometimes adopt is not opposed to the doctor's capacity to care for the other person's suffering but expressive of one form her care for the other person's suffering may take.

It seems to be the case that Stotland and Sherman are influenced by a natural scientific conception of experimental research. The idea is that we ought to make the experimental situation as restricted and controlled as possible in order to gain as clear results as possible. The problem with a reduction of context is that it becomes

unclear what it means to understand that the person is in pain.

Wittgenstein writes in *Philosophical Investigations*:

But isn't it absurd to say of a body that it has pain? —
And why does one feel an absurdity in that? In what
sense is it true that my hand does not feel pain, but I in
my hand?

What sort of issue is: Is it the *body* that feels pain?
—How is it to be decided? What makes it plausible to
say that it is *not* the body? —Well, something like this:
if someone has a pain in his hand, then the hand does
not say so (unless it writes it) and one does not comfort
the hand, but the sufferer: one looks into his face.
(Wittgenstein [1953] 2001, §286)

Wittgenstein criticizes here a conception of the human body as something that is separable from what we mean by a person. Contrary to this idea, Wittgenstein argues that it is central for what we mean by a human body, and then also for what we mean by the concept of pain, that we address the *person* who is in pain. He says “one does not comfort the hand, but the sufferer: one looks into his face”. The assumption that it is comprehensible to talk of the human body as if it was a purely physical object that we observe is reflective of the tendency to assume that sensations are a kind of inner states. The meaning of such a state is assumed to be independent of the person's larger life context. This is also an idea that Wittgenstein questions. He writes, “The concept of pain is characterized by its particular function in our life.” (Wittgenstein [1967] 1981,

§532) I have a headache that makes me grumpy. He has bad knees that makes him unable to dance. In the experiment on the other hand the persons that are subjected to the heat treatment are not affected by it in their ordinary life. After the experiment they go home and continue living as usual. The subjects who observe the treatment have also been informed that it is not dangerous. Stotland treats pain, and our understanding of it, as a contextfree momentary inner state, not as something that affects our life in various ways. There is also a difference between watching a person who has freely volunteered for a painful experiment, or seeing someone severely ill or badly injured and in pain because of it. Usually pain is not something we can freely choose to be or not to be subjected to, and it is usually not something we can walk away from when we want to. If one wants to understand how another person's pain feels, one important thing is to try to understand how this pain is a part of his life.

Let us take stock of what I have said so far. I have argued that there are several problems with Stotland and Sherman's experiment. First, the exhortations given by the experimenters seem to *steer* the participants to respond in certain ways. Stotland and Sherman do not acknowledge that ordinarily when we tell a person to use his imagination we use this expression as a specific moral reminder. We do not use the word "imagine" neutrally. That is, we do not offer the person a choice to think in any way he likes about the situation. Second, Stotland's and Sherman's experiment rests on a problematic division between seeing and imagining. They take for granted that

seeing is more “superficial” than imagining. The impression that there is a clear division between seeing and imagining rests on the fact that the context is reduced to a minimum in the experiment. This creates the impression that experiences of pain are “inner” states that we only understand by using an analogical form of imagination. Further, since the connection with real life situations of suffering is unclear this affects the participant’s ability to really understand the person’s suffering in a normal way. It becomes unclear what their compassionate or non-compassionate responses mean. This also makes it difficult to consider the results of the research.

4.4 Batson’s experimental research on empathy

Another researcher who has conducted experimental research on the relation between empathy and altruism is C. Daniel Batson. In his book *The Altruism Question* (1991) Batson reflects on whether helping has egoistic or altruistic motives. According to him, a person’s altruistic motivation to help depends on the capacity to feel empathy. Batson discusses extensively psychological experimental investigations where the relation between empathy and altruism has been studied. In these psychological experiments the results indicate, according to him, that there is a connection between empathy and the inclination to help. One of the studies he refers to is the experiment by Stotland, which I discussed above. According to Batson, Stotland’s experiment reveals in a clarifying way how the difference in perspective can affect a person’s capacity to

feel empathy. Batson concludes: "The combined evidence from these studies indicates that there is indeed an empathy-helping relationship; feeling empathy for a person in need increases the likelihood of helping to relieve that need." (Batson 1991, p. 95)

Batson has also conducted a large amount of experimental research himself in order to prove the relationship between empathy and altruism. Among other things, he constructed further experiments with Stotland's experiment as a model. Batson describes one of his experiments in the following way:

In the first two studies, female undergraduates observed a young woman named Elaine, whom they believed was receiving uncomfortable electric shocks. They were then given an unanticipated chance to help by volunteering to take the shocks in her stead. (Batson 1991, p. 113)

This experimental situation is, to a large degree, designed in a similar manner as Stotland's experiment. The test-subjects observe another person who receives painful treatment in a laboratory setting. In this sense the meaning of the experience of pain is, in a similar manner as in Stotland's experiment, made obscure by reducing the experience of pain from any normal situation such as injury or illness. My main objection to this experiment is similar to my objection to Stotland's experiment. If one reduces the normal context of how we experience pain it will also become unclear what it means to understand that the person is in pain.

However, Batson would probably not consider my critique of the experimental research method to hit the target. He is well aware of criticism directed at experimental research, suggesting that such research is too abstract or too restricted in character. According to Batson, there are two forms of empirical research methods; one that he calls the Aristotelian method and another that he calls the Galilean method. These two methods are, according to him, often taken to be in opposition to each other. According to Batson, proponents of the Aristotelian method often criticize experimental research methods for lack of "ecological validity" (Batson 1991, p. 71) Batson writes:

Aristotelian criticisms are often made of laboratory experiments in psychology today, including experiments conducted to test whether the motivation to help is altruistic or egoistic. If, for example, we set up a laboratory experiment in which we confront introductory psychology students with an opportunity to help another student in need under conditions that systematically vary the relationship between benefiting the other and benefiting themselves, we may be bombarded with Aristotelian questions like: 'Would non-students respond in the same way to the student in need?' 'Would people from another culture?' 'What if the person in need were not a student?' [...] And, most often, 'Would this need situation ever occur in real life?'

From an Aristotelian perspective, questions like these are central; they concern the historical frequency, universality, and representativeness of the

phenomenon. From a Galilean perspective such questions are quite beside the point. [...]

The Galilean scientist is concerned with a very different matter from ecological validity; he or she is testing hypothesized invariant relations among underlying constructs. [...]

From a Galilean perspective, laboratory experiments can be criticized as lacking validity only to the degree that they either (a) fail to include the variables involved in the hypothesized relation or (b) fail to exclude potential confounding variables. Whether they involve frequently observed or unusual events, naturally occurring or artificially created situations, is totally irrelevant. (Batson 1991, pp. 71-72)

Batson suggests that if we criticize the experiment for being too restricted in character, we will have missed its point. A few lines later he continues:

If we are to understand the motivation underlying [...] helpfulness, then it seems necessary to employ conditional-genetic motivational concepts such as force, goal, and conflict—concepts that refer to the dynamics that lie beneath and behind the phenotypic manifestations of helpfulness. We must adopt a Galilean approach. (Batson 1991, p. 72)

According to Batson, there is a difference between studying the frequency of a phenomenon (such as, for instance, the frequency of acts of helpfulness) or, on the other hand, studying the causal motivational factors behind helpful behaviour. He is not interested in studying

how helpfulness can vary depending on circumstances. He maintains therefore that it is irrelevant to criticize his experimental research for not having any connection with real life situations.

The question then is whether my critique of Stotland, Sherman and Batson's experiments misses the point? Partly my critique of the experiments is that they lack what Batson calls "ecological validity". However, even if I have argued that the experiments are too distanced from real life situations my aim with this criticism differs from what Batson calls the Aristotelian perspective. I have not wanted to claim that Stotland and Batson ought to study the *frequency* of helping. Rather, by pointing at the great discrepancy between the experimental situation and real life, I have suggested two main problems. First, if we reduce all normal life context we also lose the meaning of human responses. Second, the test subjects are exhorted to take certain attitudes towards the persons who are suffering. That is, the questions are leading.

Since the responses in the experiments are so far removed from ordinary circumstances of suffering, it is not clear what the responses mean. In real life suffering has real consequences, but in the experimental situations this is largely not the case. In real life we also often stand in some form of relationship to the one who suffers, but in the experimental situations there are no close relationships involved. In real life we can also consider a person responsible for helping. But there is no question of responsibility involved in the experiments. When such aspects are reduced away in the experiments it also

becomes unclear how the compassionate responses to suffering should be understood.

However, not all experiments that Batson discusses are as reductive in their character as the one described above. Batson also describes several experiments that have a more realistic character. One such experiment is the following.⁵⁰ In the experiment there are two groups of people both of which are asked to listen to a story where a young woman is in a desperate situation. The woman called Katie Banks is “a university senior whose parents had recently been killed in a tragic automobile accident. [...] Mr. Banks did not have life insurance, and Katie was struggling to take care of her surviving younger brother and sister, ages 8 and 11, while she finished her last year of college.” (Batson, Sager, Gast, Kang Runchinsky, Dawson 1997, p. 499) The first group is asked to:

Try to be as objective as possible about what has happened to the person interviewed and how it has affected his or her life. To remain objective, do not let yourself get caught up in imagining what this person has been through and how he or she feels as a result. Just try to remain detached as you listen to the broadcast. (Batson et al. 1997, p. 499)

The second group is asked to:

⁵⁰ The experiment was first conducted in 1978 by Jay S. Coke, Daniel Batson, and Katherina McDavis. Since then Batson has conducted the experiment a number of times in slightly varying forms.

Try to imagine how the person being interviewed feels about what has happened and how it has affected his or her life. Try not to concern yourself with attending to all the information presented. Just concentrate on trying to imagine how the person interviewed in the broadcast feels. (Batson et al. 1997, p. 499)

After the official experiment the test persons are asked if they would like to help the woman practically for a few hours by posting envelopes with letters for pleas for money. This suggestion to help Katie Banks was presented as a spontaneous idea by the experimenter, and as not being part of the experiment itself. The test persons also seem to believe that the request is sincere and that it is not part of the experiment.

An important difference between on the one hand Stotland and Sherman's experiment with pain and on the other hand Batson et al.'s experiment described above is that Batson et al.'s experiment is not as abstracted from real life. The story about the woman called Katie Banks sounds like something that could happen for real. In this sense it is a story that the test-subjects do understand to be a tragedy with very difficult life circumstances as a consequence in the woman's life. Also the request to post envelopes with letters for pleas for money is a request for a meaningful action in the circumstances. This also means that the test subjects' responses can more easily be seen as reflecting how we also might behave in real life. However, despite the realistic character of the experiment there remain some

problems with the experiment, which I shall describe below.

The result of the experiment was that the persons who were asked to imagine were more inclined to help while the ones who were asked to observe were not as strongly inclined to help.⁵¹ Since the experiment has a realistic character it might seem that Batson et al.'s experiment, and the results, ought to be accepted. Nussbaum (2001) comments on Batson et al.'s experimental research on empathy:

[...] there is sufficient material in the experimental reports to see that there is also a strong relationship between empathy (or, alternatively, the judgment of similar possibilities) and compassionate emotion. If empathy is not clearly necessary for compassion, it is a prominent route to it. (Nussbaum 2001, p. 332)

According to Nussbaum, Batson's experimental research indicates a "strong relationship" between empathy and compassion. I argue that the experiments are problematic.

Even if Batson et al.'s experiment is not as abstracted from human life as Stotland and Sherman's experiment, they still have basically the same problematic idea about empathy being a mental imaginative technique we can use in order to understand others. The general form of the experiment as well as the result of the experiment is also

⁵¹ For a description of the difference in the degree of helping between the low-empathy condition and the high-empathy condition see Batson et al. (1997) pp. 500-501.

basically the same as in Stotland and Sherman's experiment. The people who are asked to remain "objective" and "detached" do not respond as emotionally and "compassionately" as the people who are asked to imagine how the woman feels. According to Batson et al., this proves that our moral engagement is dependent on our using or not using a mental technique of imagination when trying to understand others. But, as in Stotland and Sherman's experiment, in this experiment, too, the researcher's instructions are not necessarily understood in a neutral way. The instructions seem to entail a certain attitude. It appears that the "objective" group is urged *not* to become engaged while the groups that are exhorted to imagine appear to be exhorted to engage morally in the other person's life. That is, even if the researchers might have intended to give certain neutral instructions it is not evident that this is how the instructions were understood by the test persons. We do not usually talk in a neutral spirit about other people's suffering, and this can also affect how we understand a researcher's instructions. This complicates how the researcher's instructions are to be understood. If the participants did not understand the instructions as requests to use a neutral cognitive technique it is also likely that it affected their responses. The results can therefore be seen as an outcome of flaws in the ways the experiment is conducted in roughly the same way as Stotland and Sherman's experiment.

Another problem is that it is unclear what sort of objective stance the test persons are actually exhorted to take. For instance, ordinarily when we talk about a judge

or a doctor being objective, we mean that the person tries to be fair and considerate. It does not mean that the judge or doctor should merely look at technical details. Trying to be fair to others also involves caring about others, not being inhuman or cold or callous. In this sense one can say that the exhortation to be “objective” means something different in the experiment than what we ordinarily mean by the word.

To sum up so far. I have tried to question experimental research that is considered to prove that there is a link between empathy and compassion. I have indicated that there are several problems with these experiments. Both in Stotland and Sherman’s experiments, and in Batson et al.’s experiments, the researchers do not seem to notice that their uses of expressions are ambiguous. Their instructions are leading, and this also reflects the results of the tests. Another problem is that the context in the experimental situations, especially in Stotland and Sherman’s experiment, is so strongly reduced that it becomes unclear how the test-subjects’ responses reflect anything about how we understand and respond to other people’s suffering in real life. In Batson et al.’s experiment with the story about Katie Banks the context is more realistic but I have claimed that the questions are still leading.

4.5 Cognitive conceptions of empathy and conceptual confusion

There are some further conceptual problems with Stotland, Sherman and Batson et al.’s experiments on empathy. The

problem is not only that the questions in the experiments are leading or that the contexts are too reduced so that the meaning of the situation becomes obscure. The experimental research also rests on the assumption that the second-person perspective (where I understand how *you* feel) is dependent on the first person perspective (where *I* feel something), when we talk of understanding. This assumption also affects Stotland's and Batson's research on the relation between empathy and compassion. The idea behind both Stotland's and Batson's explanations seem to be that I am able to care for another person by actively trying to imagine how it would feel to be this other person myself. In order for me to understand you I imagine that I have your thoughts and your feelings. In this sense the first-person perspective is considered basic for what we mean by understanding. Such a conception shapes Stotland, Sherman and Batson et al.'s experiments on empathy. Clearly it can sometimes be an important aspect of our capacity to understand another person's suffering that we have experienced something similar. But the problem arises when it is assumed that similar experiences, or the capacity to imagine oneself in the other's situation, would be a *general cognitive function or method* that enables us to understand other people's suffering.

However, Batson's and Nussbaum's conception of empathic imagination is not purely analogical. Both point out that it is an important aspect of empathy that I am aware that the suffering person's situation is not the *same* as mine. Batson notes that it is an important part of empathy that it is an other-directed emotion that is

“distinct from personal distress”, but he still describes this as a “vicarious” emotion. Nussbaum concludes that empathy “involves a participatory enactment of the situation of the sufferer, but is always combined with the awareness that one is not oneself the sufferer.” (Nussbaum 2001, p. 327) From this one can get the impression that Batson and Nussbaum actually see the difference between a first-person perspective and a second-person perspective. Still I think they do not really see the character of this difference. Even though they point out that it is an important part of empathic imagination that I am aware that the suffering person’s perspective is different from my own, they tend to describe this difference in perspective as cognitive. That is they do not see that the difference between the first-person perspective and the second-person perspective is a moral difference.

In an article, on the role of imagination in philosophy and the humanities (2002)⁵², Lars Hertzberg says that when we talk about someone being able to imagine another person’s suffering we use the word “imagine” in a moral sense, we say something about his being attentive to others, or fair-minded, of his being neither sentimental nor cynical. (Hertzberg 2002, pp. 8-9) One is not here talking about what goes on inside his mind or what he feels. One describes his ways of *being* with others, his ways of acting and paying attention to others in difficult situations.

⁵² Hertzberg’s paper is written in Swedish. The title is: ‘Om inlevelsens roll i filosofi och humaniora’. In English the title means: On the role of imagination in philosophy and in the humanities.

Hertzberg gives the example of a man who is severely ill and tells a doctor that he wants to die and that his treatment should be stopped. A doctor can feel that it is difficult to know how to relate to the man's words. This does not mean that there is one right answer to be found if we managed to see through him. It is not even certain that the person himself actually knows whether he really means what he is saying. Hertzberg's point is that it is because the situation is of such a serious character that one may become uncertain how to relate to the suffering person's words. The reason we do not constantly find it difficult to know what other people feel or what they mean by their words is not a matter of our usually having more information about others but rather of the situations being harmless and ordinary. The way one attaches importance to the person's words as well as the way one can be uncertain about how to take his words, is an expression of one's sense of the serious character of the situation. This reaction in itself is a form of moral awareness of the other, which also shows in how one acts, in how one listens to the person, how one might be sensitive to his gestures etc. The way questions and uncertainty about the meaning of the other person's words can come in here is in itself an expression of a compassionate attitude towards the suffering man.

In *Totality and Infinity* ([1961] 1969) Emmanuel Levinas writes: "The face in its nakedness as a face presents to me the destitution of the poor one and the stranger [...]" (Levinas [1961] 1969, p. 213). A bit further he continues; "To hear his destitution which cries out for justice is not to

represent an image to oneself, but is to posit oneself as responsible [...]” (Levinas [1961] 1969, p. 215). A bit further Levinas continues:

The Other who dominates me in his transcendence is thus the stranger, the widow, and the orphan, to whom I am obligated. These differences between the Other and me do not depend on different ‘properties’ that would be inherent in the ‘I,’ on the one hand, and, on the other hand, in the Other, nor on different psychological dispositions which their minds would take on from the encounter. (Levinas [1961] 1969, p. 215)

Levinas here talks of a difference in perspective that does *not* lie in the different properties of the two individuals’ minds, but in the other person appealing to my *responsibility*, I am *obligated* to help the other. That is, according to Levinas, my awareness of there being a difference in perspective when I am confronted by “the orphan or the widow” is *in itself* expressive of a sense for the seriousness of the other’s life situation; it is expressive of a sense of responsibility of caring for the other.

Hertzberg and Levinas suggest two aspects of what we mean when we talk of understanding another person’s suffering. First, Levinas suggests that our awareness of another person’s suffering is integral to the fact that we are responsible for the other and thus that we respond to the other, by for instance trying to help him or her. Second, Hertzberg suggests that questions that may arise about how the suffering person feels are in themselves expressive of our sense of care and responsibility for the other person.

In this sense questions about the other person's perspective are not neutrally cognitive but moral.

However, it can often be the case that we do not really know how another person feels or thinks about her or his situation. In this sense our failure to understand another person's suffering can sometimes partly be of epistemological character. However, Batson and Nussbaum consider such differences in perspective as a general cognitive problem that ought to be solved by the use of a general cognitive method of imagination. I think it would be better to consider such differences in perspective as specific to certain life situations and certain relationships rather than as a general cognitive epistemological problem of other minds. Further, as Levinas and Hertzberg suggest, the way we can feel that the suffering person has a different perspective, the way we can worry and feel that we do not know how he feels, is often in itself expressive of our care for the other.

Batson and Nussbaum's emphasis on imagination as a cognitive tool that enables us to understand the suffering person is also connected with a tendency not to acknowledge that it is a central aspect of compassion that we *talk* with the suffering person. Compassion has a mutual character in how one acknowledges one's presence to the other by talking. One might, for instance, say worriedly to a sick friend "How are you feeling?" This way of talking is one of the more central ways of showing compassion towards another person; it is a way of acknowledging the other person's suffering, a way of expressing respect, warmth and of caring for the other.

One *shows* that one cares about the other person, one *lets him know*. Bodily expressions can also in this sense, as Wittgenstein suggests⁵³, be of great importance as expressions of compassion; that one looks into another person's eyes, that one does not leave a suffering person alone, that one holds the person's hand. Knowledge comes into compassion very much as something one shares with the other, in one's attentiveness towards the other.

The central form of the relation between compassion and knowledge here is not that we first need to know exactly what the other person is feeling, which again makes us care for the other person, but that our care for the other can show in that we may worry about how the person feels and that we also then may ask the person how he or she feels. The way we consider the suffering person's first-person perspective to have special significance, including the way we may feel that we do not know how he feels unless we *ask* him, is in itself expressive of our care and respect for the person. The concept of knowledge is, in this sense, relational and moral.

Compassion also often takes the form of a respectful acknowledgement that one does *not* have the authority to say how things are for the other. C.S. Lewis writes in *Letters to Malcolm* ([1964] 1991)⁵⁴ about the difficulty of his grief after having lost his wife. "You wrote; 'I know I'm outside. My voice can hardly reach you.' And that was one reason why your letter was more like the real grasp of a

⁵³ See the earlier quote from Wittgenstein, ([1953] 2001, §286).

⁵⁴ Malcolm is generally considered to be a fictional character.

real hand than any other I got.” (Lewis [1964] 1991, p. 41) Lewis illustrates here how a friend’s expression of compassion can show in a respectful and honest acknowledgement of how grief can be a lonely struggle.

4.6 Imagination, involvement and responsibility

I have argued so far that the idea that empathic imagination causes compassion is reflective of a conceptual confusion between the first-person perspective and the second-person perspective. I have also claimed that there is a tendency not to see that this difference is not cognitive but relational and moral.

However, even if the concept of empathy entails conceptual confusions there are cases in real life that can seem to reflect the theories on empathy. In the following I shall quote some passages from Gösta Karf’s book *En junisöndag kvart över tolv*⁵⁵. In this book Gösta Karf tells about his grief after his five-year old son Markus is killed in a fire. Markus had been playing alone in the family’s car when the car suddenly, without any obvious reason, started to burn. Markus didn’t manage to get out of the car and was killed in the fire. Several years after the accident Gösta Karf wrote a book about the family’s grief. The book contains many descriptions of how people acknowledge or fail to acknowledge the family’s grief. Karf writes as follows about the moments when Markus is brought out of the burning car.

⁵⁵ Karf’s book is written in Swedish. In English the title means: A Sunday in June a quarter past twelve.

Suddenly something is about to happen. The rescue personnel are preparing something. [...] A blanket is placed on the ground behind the bushes, at a little distance from the car. The curious onlookers that have come to satiate their hunger for sensation are driven away. The firemen are standing respectfully still, some with their gaze lowered to the ground. I cannot see what is happening inside the car, but I begin to realize what is going on. Then the fireman comes out with little Markus in his arms. Carefully and tenderly he is placed on the blanket that has been brought from the burned car's boot. One of the firemen places one of Markus' teddy bears beside him. Markus' body is taken to the ambulance. A fireman sits beside him and then the ambulance drives to the hospital. The atmosphere is tense to the breaking point. I can see that some of the firemen have a difficult time struggling with their tears. [...] One of the firemen does not manage to hide his curiosity but asks Marita: 'Was it a girl or a boy?' Then we realize how badly burnt Marcus was. This insight was given us too soon. It hurts incredibly to be informed of the fact that the person one loves so much and who one only a few hours ago held in one's arms, now is so severely injured that one cannot see whether it is a girl or a boy. For the relatives of a severely injured person it can be of essential importance to be given the chance to live with the illusion and belief that the victim is perhaps not so deformed, even if deep inside one suspects the opposite. (Karf 1999, pp. 19-20, my translation)

One can here see a variety of responses. The firemen struggle with their tears. They also handle Markus' body

very tenderly (even though it is clear that he is already dead). They place him carefully on a blanket and one fireman puts Markus' teddy bear beside him. The firemen act with compassion. However, a bit later in the quote above, one of the firemen asks an insensitive question. He asks the parents whether the dead child was a boy or a girl. The fireman surely did not want to offend these parents. What was it that happened then? One could say that the fireman's question was expressive of an inability to imagine the parents' perspective; i.e. an inability to imagine the effect of his question. However, it might be that he actually tried to say something kind. Sometimes we are clumsy when we try to be kind. Perhaps the fireman's question was partly expressive of inexperience. Still one could say that he lacked a capacity to imagine the effect of his words. Here is another description by Karf.

The foreman of the fire squad comes and offers us his condolences. From now on none of us is left alone, not for a second. The foreman keeps us all under his watching eye. With every step my wife takes he walks beside her and he also follows me with his eyes. It feels safe that someone at last takes responsibility of the situation. (Karf 1999, p. 21, my translation)

The foreman's behaviour here brings to mind what Levinas talks about when he says that the other person's suffering poses me as responsible. The foreman takes responsibility for the situation. Karf continues with the following description:

Our ambulance stops in front of the entrance to the emergency ward. The doctor on duty and a nurse stand ready to meet us. The doctor seems insecure and nervous. When we step out of the ambulance he comes to meet us and says briefly and formally: 'I looked at the boy but there was nothing to do.' He then leaves quickly without saying anything more. Then we are led to a room that is meant for patients who need to lie down. We are asked to wait. 'the doctor will come soon.' the nurse says. [...] The doctor approaches us quickly. Now we shall finally get time to talk, I think and I put all my trust and responsibility on the doctor. He continues to give his brief information: 'In case you will be transferred to the hospital I will take care of you, but then you must first get a referral.' Again he leaves as quickly as he came. Speechless I stand and look when he leaves. He appears to completely have lost his composure, and I can barely believe what I see and experience. [...] I do not know whether I should laugh or cry. (Karf 1999, pp. 22-23, my translation)

Here Karf describes the doctor as if he was completely unable to acknowledge their grief. The doctor's behaviour is in some sense reminiscent of the "technical" or "objective" perspective that some of the experimental subjects are asked to take in Stotland and Sherman's and Batson et al.'s experiments. One might also say that the doctor seems incapable of imagining the grieving parents' perspective. Such a description can tempt one to think that there is after all something correct in Stotland and Sherman's and Batson et al.'s experiments. The thought would then be that if the doctor had been capable of

imagining the parents' grief, he would have been more compassionate in his behaviour. However, such an explanation appears odd. Nothing could be more evident in the situation than the fact that the parents are struck by grief. Another way to put it could be to say that it is as if the doctor is scared and hides his inability to approach the grieving parents under the clinical instructions. The doctor is not neutral in his clinical way of behaving. On the contrary, his way of sticking to purely clinical instructions is highly emotional. From Stotland and Sherman's and Batson et al.'s experiments one might get the impression that a capacity to observe technical details is something neutral, something that does not express our involvement with the other person. I have maintained that this is not the case in the example above. A bit further on Karf continues with the following description:

News about the accident has spread fast and has reached out to the circle of friends and acquaintances, who one after another come to express their condolences and to offer some words of consolation. I can see how incredibly difficult it is for many of them to find any words that can express what they feel. The words are not important. No words can shatter or even ease our grief but their presence gives us comfort. One of the visitors tries to lighten up the atmosphere and to break the silence by saying something funny. Someone shouts several times: 'Markus is all right now, Markus is all right!' The words are well meant, but regardless they remain incomprehensible. How could he be all right when he cannot be together with his family, I

think. How can he be all right when we are suffering? A few hours ago he was all right, when he was lying in Marita's and my arms with his cherished milk bottle.

Where are the words that have meaning and depth in the situation, I wonder. What words would I most like to hear? That question is not easy to answer. (Karf 1999, pp. 33-34, my translation)

When several of the friends of the family visit they cannot find any words to say. Does this mean that they cannot imagine the Karfs' grief? Several of the guests also behave a bit awkwardly even if they want to offer consolation; they try to lighten up the atmosphere, they say things that Gösta Karf experiences as incomprehensible. At the same time he notes that the words do not matter. It does not matter that some expressions are somewhat awkward or that some even sound unintelligible, but the friends' presence is comforting. Karf also notes that he himself does not know what words of consolation he would like to hear. He does not know what words could help and console. Some time later Gösta and Marita Karf have to visit the funeral parlour to buy a grave.

A woman asks us kindly to sit. From her appearance one can see that she knows why we have come. [...] It is quiet for a long time, finally I say that we want to buy a grave. The woman looks at us and then she turns her eyes down towards the desk and begins to cry. Embarrassed she apologizes for her emotional outburst. For the first time we experience a person with a strong capacity to imagine, who dares to show what she feels. (Karf 1999, p. 45, my translation)

This description of the woman's reaction seems to resemble the situations in Stotland and Sherman's and Batson et al.'s experiments where the subjects become emotionally moved when imagining the other person's suffering. It is also clear from Karf's description that they deeply appreciate the woman's reaction. However, there is still a difference between the woman's reaction and the suggestion that we use an empathic method of imagination in order to feel compassion. In the case above the woman has not decided to imagine how the parents feel. Rather, when suddenly meeting the grieving parents the child's death and the parents' grief becomes so present to her that she cannot help crying. The woman's reaction reflects Levinas' thoughts about the ethical meaning of the other person's face. It is when the woman *meets* the parents that she begins to cry. At the same time her reaction is not merely a spontaneous emotional outburst, Karf describes her reaction as courageous and honest. However, how a person is moved by another person's suffering can also show in other ways. Consider the following passage in which Karf describes how they meet the owner of the funeral parlour.

The owner of the funeral parlour asks us to join him in his office. To me it feels better like this. I do not have to face all the coffins and urns. He takes the initiative in the conversation and it feels relaxed. He practices his profession with great dignity. It is unusual to meet a man with such good intuition. He observes us carefully without being intrusive. I get the impression that he is aware that he should be very cautious and gentle. He

asks carefully: 'Do you know where Markus is now?' Since no one has informed us of where he has been taken and where and when the autopsy will take place, we get a precise description of the formalities. The owner of the funeral parlour is well informed of everything that happens now [...]. His considerate and dignified manner is warming. Every time Markus is mentioned in some context he says his name instead of 'the boy', 'the deceased' or something else impersonal and cold. [...] There is a warmth in his whole demeanour and we get unlimited time to talk. All the descriptions and practical advice we get are well thought through and useful. In passing he asks how tall Markus is. I realise that he needs this information for the coffin. [...] I have to admit I feel some discomfort about the requested information about Markus' length. The question is unavoidable, but if the conversation had started with that question our presence at the funeral parlour would have had a very sudden ending. (Karf 1999, pp. 48-49, my translation)

The owner of the funeral parlour has obviously had much experience in meeting people in grief. This has apparently made him very considerate and sensitive. The fact that he gives Gösta and Marita Karf much time to talk, and the fact that he uses Markus' name when he talks about him, are of great importance for them. Here one could say that the man has learned to imagine the grieving person's perspective. He is also very good at informing the Karfs about practical details. In contrast to the previous example with the doctor whose manner of giving brief practical information was expressive of his own fear, the owner of

the funeral parlour's detailed practical knowledge about the procedures of the autopsy is expressive of his care and respect for the grieving parents. In this sense there is not here the kind of division between a capacity to observe technical details and on the other hand a capacity to imagine how the suffering person feels, that Stotland and Sherman's and Batson et al.'s experiments presuppose. The undertaker is clearly skilled at sensing the importance of certain details. His sense for details shows, for instance in how he mentions Markus by name and not in an impersonal sense as "the boy" or as the "deceased". It also shows in his careful way of waiting with the question about Markus' length. Karf also says that the man has an unusually good intuition. But does this mean that he has learned to use a general mental technique of imagination? It would, I think, be problematic to say this. Surely the owner of the funeral parlour has reflected much on what it means to meet people in grief. But the way he has reflected and the fact that he senses that certain details are important, is expressive of his character as well as of his professionalism. His whole demeanour is expressive of his respect and care for Gösta and Marita Karf's grief.

According to Hertzberg (2007) philosophers tend to take for granted that the word "think" refers to a neutral cognitive process. Hertzberg claims that philosophers have tended to make a dichotomy between thinking as an intellectual process and a person's character.

The grammar of the verb 'to think' leads us to suppose that our ability to perform in a thoughtful manner must

be based on some activity that we carry out, some effort we undertake or some process that takes place within us. [...] we assume that this is a neutral, inert, mechanical, impersonal procedure by which conclusions are deductively or inductively derived from premises independently supplied. It is taken to be 'impersonal' in the sense that it is routinely accessible and yields the same results to anyone willing to make the effort [...] In fact, some of the most emphatic uses of that word [think] are ones that involve a reflection on a person's character as when we blame someone for thinking only of himself, or when we praise a person's thoughtfulness, say, in choosing the perfect gift for someone else's birthday. (Hertzberg 2007, pp. 68-69)

What details the owner of the funeral parlour sees as important, how he reflects, is in itself expressive of his care and respect for the Karfs as well as of his capacity to assume responsibility in the situation. When the Karfs leave the funeral parlour and walk home they see an acquaintance:

A few blocks further away we see another acquaintance who works in the healthcare. She flees quickly over to the other side of the street happily presuming that she has not been noticed. (Karf 1999, p. 50, my translation)

Is the woman here unable to imagine the Karfs' grief? Or is it rather the case that she is so much aware of their grief that she becomes afraid to confront them. In several of Karf's descriptions one can see how people become afraid

and avoid the Karfs. It is as if people become unable to meet them because of their enormous loss. These are not reactions of callousness, nor are they expressive of a cognitive inability to imagine what has happened. However, Karf also describes several situations where people do acknowledge their grief and show compassion.

To sum up so far, I have claimed that it is problematic to think of compassion as if it was dependent on a general method of imagination. I have also questioned the assumption that our difficulties to understand other people's suffering consists in a general cognitive difficulty to imagine the suffering person's perspective. However, by this I have not meant to say that we have no difficulties in imagining a suffering person's perspective. Severe suffering can affect a person's life so fundamentally that it can, for other people, be very difficult to really comprehend and acknowledge the character of such suffering. This can make us clumsy when meeting such a person. We might, for instance, as in the case with the fireman mentioned earlier, say things that would sound all right in an ordinary situation but that can be devastating for the suffering person. However, these difficulties to imagine the consequences of one's words cannot be solved by the use of a general cognitive method of imagination. The difference between the persons who acknowledge the Karfs' grief and the ones who do not is not a difference in a general cognitive capacity, it is a difference in how they handle certain situations, how they assume or do not assume responsibility, how they are, or are not, careful with their words, how they are experienced or not etc.

4.7 Solipsism, body-mind dualism and experiences of suffering

Earlier I criticized especially Stotland and Sherman's experiment for being problematic in that it presupposes a body-mind dualism. This does, however, not mean that body-mind dualism is merely a philosophical hang-up. Body-mind dualism is often referred to as having its origins in René Descartes' philosophy. However, as I mentioned in the introduction, one could also say that body-mind dualism partly has its roots in medical science, by which Descartes was strongly influenced. In medical science the body becomes a physical object to be observed. Similarly, in medical science pain becomes a state one can observe and test in a clinic. It appears to be such a medical and anatomical conception of the body as something to observe, and also such a medical conception of pain as a momentary state of a certain magnitude, that Stotland and Sherman's experiment is based on. The problem arises when such medical ways of looking are dislodged from the context of ordinary human life, a life where the pain bothers us, where it can go on for days or weeks, where it is something we can't get rid of, where it affects our eating, sleeping, walking, talking etc. The problem appears when the medical perspective is thought to be "purer" or more "basic" than the fact that pain bothers us in real life in many ways. From this idea of the pain in a medical situation being "purer" than the pain in real life it is not a big step to beginning to think of pain as a private sensation.

However, pain also has other traits that in a sense can explain why it is considered to be something private or inner. In *The Absent Body* (1990) Drew Leder describes several patterns in experiences of pain. He points out that pain makes us self-aware in a sense in which we are usually not. But this self-awareness can at the same time constrict our ability to think, reflect and sense ourselves as having a life. The body becomes the center of our focus, while the ordinary life surrounding disappears.

With chronic suffering there is nowhere to go, nothing to do, no escape. Space loses its normal directionality as the world ceases to be the locus of purposeful action.

Physical suffering constricts not only the spatial but the temporal sphere. As it pulls us back to the *here*, so severe pain summons us to the *now*. [...]

This temporal constriction is characteristic of chronic pain as well. While the body in well-being can explore the far reaches of time through memory and imagination, such possibilities constrict when we are in pain. (Leder 1990, p. 76)

Pain as an experience can overwhelm us so that it alienates us from life. This Leder sees as partly explaining why we tend to think of body and mind as separate entities. Strong pain grabs our consciousness in a way that makes us unable to engage in ordinary life, as well as to concentrate on things, to reflect, to think, to remember, and to pay attention to others. Pain constricts our being into a forced awareness of the body. Karen Fiser (1986) writes:

Pain sometimes suffuses the person's world, so everything is stained with it. Sometimes the world is charged with energy and well-being, one's eyes sparkle and one's body feels alive. And then, on a different day, the world drops away. (Fiser 1986, p. 8)

In this sense one can say that there are two kinds of body-mind dualism; one that rests on philosophical confusion concerning interpersonal understanding (and that seems to be influenced by a conception of the human body that is common in modern medical science), and another that reflects certain experiences in real life of suffering. Leder writes:

Thus, the sense of the body as an alien thing does not arise solely in the objectifications of the modern physician. Prior to visiting the doctor's office, the pain and disability of the patient have already laid the groundwork for a distanced perspective. Plügge,⁵⁶ the German physician and philosopher, has discussed this phenomenon eloquently. He argues that the sheer 'thinglike' nature of the body, as reified in Cartesian metaphysics, first surfaces in life-world experiences of effort, fatigue, disease, and the like. (Leder 1990, p. 77)

Pain can also isolate us from others. In his book *Must we Mean What we Say?* ([1969] 2002)⁵⁷ Stanley Cavell draws a parallel between solipsism and the loneliness one can experience in severe suffering. Even if Cavell does not

⁵⁶ Leder refers here to Plügge (1967).

⁵⁷ Chapter "Knowing and Acknowledging" in Cavell ([1969] 2002).

think that all our feelings are private inner experiences such that no one else can know how they feel, he still sees a feature of this thought that is important to take into account when one talks about suffering. According to Cavell, there is a character of loneliness and exposedness in serious suffering that makes us talk of it as private and as something others cannot understand. It is true that compassion is an important expression of our understanding of another person's suffering, but when suffering is really difficult we often can't help the one who suffers, or our help cannot give much relief. This, according to Cavell, is one explanation why we can talk of pain as something inner and private. Tito Colliander writes in his book *Samtal med smärtan* (1956),⁵⁸ about the pain he has when he is suffering from a severe disease.

Sometimes you look up from a hole in the ice, you stick your head up and take a deep breath and smile towards the sun. But a moment later you have again been swallowed by suffocating water. [...] And more and more clearly you realize that no human being, however close she may be to you, can carry any of the pain in your limbs. It is totally and fully only your own. (Colliander 1956, pp. 17-18, my translation)

Colliander describes his intense pain as an experience of loneliness and alienation from life, as if he were suffocating, and as something others can't help him with. It is

⁵⁸ Colliander's book is written in Swedish. In English the title means: Conversations with pain.

important to see that this sense of the pain being something “fully your own” is a description of the unbearable character of the pain; it is not a philosophical statement about the private character of all our sensations. A difference between the philosophical idea of pain as inner and the loneliness in suffering is that the philosophical conception of pain as inner is a calm conclusion; it is not an expression of a desperate experience of helplessness. From the philosophical point of view, the reflections on pain as something inner is often made with the view that *all* our feelings are private inner states, it can equally much concern the fact that my foot itches or that I like the taste of chocolate. The private character of pain then becomes a general problem of knowledge that concerns all our feelings rather than the sense of isolation being seen as integral to suffering. Leder also points out that the whole idea of our feelings all having a similar character of privacy distorts what it can mean to feel something. Many of our pleasurable feelings *unite* us with others and gain their meaning through our being together with others.

Pleasures, as more tied to a common world, also tend to maintain our intentional links with other people. We feast and drink with friends, making of our enjoyment a common bond. It is our means of connection, not, as Updike⁵⁹ writes of pain, a ‘filthy window’ interposed between us. The primal image of pleasure, that of the infant feeding, depends upon a caretaker’s presence. In

⁵⁹ Leder refers to Updike (1983).

adulthood many pleasures, such as the sexual, are still secured primarily with and through others. [...]

Thus, as Buytendijk⁶⁰ discusses, pleasure and the happiness with which it is often accompanied is naturally 'expansive'. We fill our bodies with what they lack, open up to the stream of the world, reach out to others. In contrast, pain tends to induce self-reflection and isolation. It effects a *spatiotemporal constriction*. (Leder 1990, p. 75)

The problem is that intense pain as an experience of separateness and alienation from life is, in philosophical discussions, easily misconstrued as a question of all our sensations being inner in the sense that they can only be directly and fully understood from a first-person perspective.

Another difference between philosophical conceptions of pain as inner (in the sense that all our sensations are considered to be something inner) and how suffering can be a lonely experience is that in philosophy pain is reduced to a context free and timeless inner state. But in real life the isolating character of suffering can usually only be understood by seeing the whole life context. Important then, as I have argued before, is that strong pain is often connected with serious illness, disability and death. Not only is pain frightening because of how it feels (which already means that it can be immensely frightening), but it is also frightening because it often means death. Being in difficult pain also often means that you have to enter

⁶⁰ Leder refers to Buytendijk (1962).

places that you normally do not enter, such as a hospital. Such aspects as having to spend weeks or months in hospital can have an alienating effect on the person who is ill. On the one hand you are seriously ill and you are scared and worried about your future, perhaps you also know you are dying. Often you are also so tired that you do not have the strength to pay attention to ordinary daily matters, not even the strength to try to get distracted by looking at TV, or the strength to plan, reflect on or worry about ordinary matters (bills to pay, work problems, the family etc.). In a hospital you are also in a kind of non-place, together with people you do not know, and also usually not left alone. Further, even though the hospital staff is often kind and of tremendous help these are still people you do not know very well. Lying in hospital stops you from living a normal life, stops you from sharing daily joys and worries with your family and friends. One loses the spontaneous daily contact and daily routines with others that largely form and give meaning to our life, but that also enable us to think of other things than ourselves. There is, in a sense, a loss of an ordinary engaged perspective and a forced awareness of one's own state. In such a situation it can be of immense importance that close ones visit and call and tell about ordinary things that are going on; it can bring back a sense of being part of life, and it can momentarily make one forget oneself.

However, even if strong, long-lasting pain or illness can be a very lonely experience, it does not have to mean that one can't tell others how it feels, or that others can't see or understand how it feels, or that one necessarily thinks

others are no support in one's suffering. Rather, for the one who has to witness another person's strong pain it can be obvious that the pain is a lonely experience, and one can show this by, for instance, holding the suffering person's hand and by not leaving him or her alone. At the same time, however, the one who is in pain can sometimes experience a gulf between his life and the life of people who are well. This sense of separateness can partly be due to the severity of the pain itself, but it can also have to do with a sense of inability to really explain or talk about one's pain. Fiser (1986) talks about how a patient sometimes can feel bewildered about the pain, not knowing really how to describe it to the doctor.

In the article "Night" (2010) Tony Judt writes about the illness ALS, which he was suffering from. This is a difficult illness that slowly makes a person completely paralyzed and eventually leads to death. Tony Judt suffered from ALS for two years and finally died from it in 2010. For Judt the long nights were most difficult to bear.

In the early stages of my disease the temptation to call out for help was almost irresistible: every muscle felt in need of movement, every inch of skin itched, my bladder found mysterious ways to refill itself in the night and thus require relief, and in general I felt a desperate need for the reassurance of light, company, and the simple comforts of human intercourse. (Judt 2010, p. 3)

The sense of isolation, especially during night, was very strong. Judt further notes:

[...] it is hard to resist the thought that even the best-meaning and most generously thoughtful friend or relative cannot hope to understand the sense of isolation and imprisonment that this disease imposes upon its victims. (Judt 2010, p. 4)

Judt's reflection above might be understood as if he gave a cognitive statement about the privacy of our feelings. But this is, I think, not what he is saying. On the contrary he is expressing by his words the extreme sense of isolation that he experiences in his illness. It was apparently a form of isolation that he felt was difficult for others to truly acknowledge and comprehend, despite the fact that people were thoughtful and considerate.

In another article, "Tony Judt: A Final Victory" (2012) written by Judt's wife Jennifer A. Homans, she also tells about her husband's life when he was suffering from ALS. She describes how Judt's illness made them both live in a kind of "bubble", separated from ordinary life.

At some point—it is hard to say exactly when, but it was about the time he began *Thinking the Twentieth Century*—we entered what we came to call the bubble. The bubble was a closed world, an alternate reality, a place that we lived in and peered out of. It had walls—transparent, filmy walls—but they were like one-way mirrors: we could see out, but no one could really see in, or at least that is how it felt from the inside. We knew our world was strange and apart, governed by the rules of illness and dying rather than the rules of life. I could pierce through, sometimes, by taking a

walk and seeing the sky, but Tony could not—and increasingly would not. (Homans 2012, p. 4)

Homans' description here of their living in a "bubble" is one way of describing a sense of estrangement from ordinary life. At the same time this "bubble" is something they share; they are *both* in this bubble. Homans describes the "bubble" as both a refuge and a prison. "He took grim refuge in his study, his sickroom, his closed, safe prison-cocoon that would house his deteriorating body and entrapped mind." (Homans 2012, p. 4) In this bubble Judt found comfort in close ones and friends, but at the same time he also felt he had lost his old self.

His private life at home and with friends was his greatest comfort but it was also deeply sad: he couldn't be the things he wanted to be and he was haunted and humiliated by his 'old' self—what he called 'the old Tony,' who was lost to him forever. (Homans 2012, p. 4)

Judt's illness did not only distance him from the ordinary life of others but also made him unable to be the person he had once been. So he found even meeting close friends saddening. However, for Judt intellectual work had always been central in his life, and it was also through such work that he was able occasionally to be himself again and find some relief from his illness. One way to get away from the illness was the internet. By emailing others who could not see him, and people who did not know him, he was momentarily able to get away from the illness.

With the help of his family and friends and especially his extraordinary assistant, Eugene Rusyn, who had a way of effacing himself and could type at the speed of thought and speech, Tony could sit at the computer and we could act as his hands, typing his words and opening his view electronically out onto the world. And so he took more and more writing, more and more e-mail and electronic interviews; anything where people could hear or read but not see. (Homans 2012, p. 4)

Through the help of others with typing among other things, Judt was able to leave his embodied self and engage in conversations so that his illness was momentarily lost in an unseen background. Through the help of others he was able to gain his former intellectual independence. Homans' description of Judt's situation reveals how much our experience of illness can be dependent on how other people acknowledge us. Both Judt's awareness of his own illness, his depression over not being able to be "his old self", as well as his capacity to occasionally forget his illness was largely formed through his relationship to other people. This also goes against the idea that pain is something that we have access to from a first-person perspective where other people have no role for the character or meaning of this first-person perspective. Rather, it seems that Judt's first-person perspective in many ways was integral to his relationship to other people.

Another thread back to life was his work on the book.

Memory was Tony's only certainty and he clung to it as a lifeline. It was the thing the disease could not take from him. It was another way out of the bubble and the only form of independence he had, and kept, to the very end. (Homans 2012, p. 4)

Judt worked on the book with the help of his friend Timothy Snyder.

For Tony the incentive behind the book—and it had to be a powerful one to overcome the discomfort and depression that were his constant companions—was primarily intellectual, a matter of clarification. Tim knew this, and when their dialogue worked, as it usually did, Tony was transformed. Sick Tony, frustrated and anguished Tony, unable to eat or scratch or breathe properly, his body aching from inactivity, was able, with Tim and through sheer mental and physical exertion, to find some relief and exhilaration in the life of the mind. (Homans 2012, p. 4)

Judt's illness was of a very severe type, and there are of course many different kinds of illnesses and people experience illness in many different ways. However, Homan's description of Judt's illness suggests that some aspects of the way we experience pain and illness in real life can seem to resemble the philosophical temptation to think of our experiences as solipsistic. From the perspective that strong pain and illness can alienate us from our capacity to think, as well as also alienate us from our own body, the thought of the human being as consisting of two parts, mind and body can seem

understandable. Likewise, from the fact that illness and pain can isolate us from life and from others, the idea of sensations as inner and private can seem understandable. But that these philosophical perspectives can seem understandable if one thinks of illness, does not mean that pain and illness prove the philosophical body-mind dualistic thinking or solipsism to be correct. The philosophical solipsistic and dualistic perspectives get problematic when the human life context is forgotten and pain is thought of as exemplifying any kind of “inner sensation” in itself and thus also as exemplifying something “inner” in the sense of something unattainable that we can only understand indirectly by analogical imagination. Such assumptions are reflected in the experiments that I discussed earlier in this chapter.

From Judt’s and Homans’ descriptions above one also sees that the experience of illness is constantly entwined with our relationships to other people. In this sense, even if experiences of illness sometimes have a dualistic or a solipsistic character, this ought to be understood as integral to how illness can affect a person’s experience of himself as well as his relationships to other people. The dualism and solipsism are not here something that can be understood in a general cognitive sense but must be seen as integral to the experience of the illness.

In experiences of violence one can also see dualistic and solipsistic traits while these experiences cannot be understood as dualistic or solipsistic in a general, philosophical and epistemological, sense. The experiences reflect the person’s traumatized relationship to other

people as well as to him or herself. In *Aftermath, Violence and the Remaking of a Self* (2002) Susan J. Brison, a victim of rape and attempted murder, discusses what it means to experience violence. She writes:

When the trauma is of human origin and is intentionally inflicted [...] it not only shatters one's fundamental assumptions about the world and one's safety in it but also severs the sustaining connection between the self and the rest of humanity. [...] the self exists fundamentally in relation to others. (Brison 2002, p. 40)

According to Brison, a victim of violence largely loses his or her former sense of being alive; something which affects the victim's sense of thinking, remembering, feeling joy or grief. She also comments on how the experience of violence affected her sense of self.

I was no longer the same person I had been before the assault, and one of the ways in which I seemed changed was that I had a different relationship with my body. My body was now perceived as an enemy [...] as a site of increased vulnerability. But rejecting the body and returning to the life of the mind was not an option, since body and mind had become nearly indistinguishable. My mental state (typically, depression) felt physiological, like lead in my veins, whereas my physical state (frequently one of incapacitation by fear and anxiety) was the incarnation of a cognitive and emotional paralysis resulting from

shattered assumptions about my safety in the world.
(Brison 2002, p. 44)

Victims of violence often suffer from vivid nightmares and from various kinds of over-sensitivity to sounds or to certain situations. Fear of attack might make the person unable to go outside alone for walks. Fear as well as physical injury can affect the victim's ability to continue the job she or he had before, thus isolating the victim from former work mates. The person might also be unable to stand any sort of physical touch even by loved ones. In this sense experiences of violence create boundaries between the victim and the victim's former relationship to other people, including close ones. Brison also recounts how she after the assault experienced a loss of language.

After my assault, I had frequently had trouble speaking. I lost my voice, literally, when I lost my ability to continue my life's narrative. I was never entirely mute, but I often had bouts of what a friend labeled 'fractured speech,' in which I stuttered and stammered, unable to string together a simple sentence without the words scattering like a broken necklace. [...] For about a year after the assault, I rarely, if ever, spoke in smoothly flowing sentences. (Brison 2002, p. 114)

Various aspects of shame may also be involved in experiences of violence. This makes it difficult for the victim to talk about his or her experiences, but it also makes it difficult for other people to ask. Brison writes:

During the first few months after my assault, my close friends, my sister, and my parents were supportive, but most of the aunts, uncles, cousins, and friends of the family notified by my parents almost immediately after the attack didn't phone, write, or even send a get well card, in spite of my extended hospital stay. These are all caring, decent people who would have sent wishes for a speedy recovery if I'd had, say, an appendectomy. [...] In the case of rape, the intersection of multiple taboos—against talking openly about trauma, about violence, about sex—causes conversational gridlock, paralyzing the would-be supporter. [...] When, on entering the angry phase of my recovery period, I railed at my parents: 'Why haven't my relatives called or written? Why hasn't my own brother phoned?' They replied, 'They all expressed their concern to us, but they didn't want to remind you of what happened.' Didn't they realize I thought about the attack every minute of every day and that their inability to respond made me feel as though I had, in fact, died and no one had bothered to come for the funeral? (Brison 2002, pp. 12-13)

The friends' and relatives' inability to understand Brison's situation did not have to do with a lack of information about what had happened; it was more a question of a misdirected concern as well as of an inability to approach her. These reactions were connected with the fact that she was the victim of a sexual assault. As she concludes, had she been in hospital because of an appendectomy, many of her relatives and friends would have expressed their concern to her. What made the lack of response especially

hard to bear was that it was her close relatives and good friends that failed to respond; people she had had confidence in and whom she liked and knew well. One could think that this was a situation where empathic imagination was needed. But it was not so much imagination as the courage to be honest, to visit, talk with and listen that was lacking. Brison is not saying that she was disappointed with her relatives because they lacked the ability to imagine what it might mean to be the victim of a sexual assault. It was not an expert capacity to imagine details that she longed for, but simply the presence of close ones and an honest acknowledgement of what had happened to her.⁶¹

Brison further reflects on the importance of being listened to and of being able to talk about one's experiences. She notes: "*Saying* something about a traumatic memory *does* something to it." (Brison 2002, p. 56)

By constructing and telling a narrative of the trauma endured, and with the help of understanding listeners,

⁶¹ The stigmatizing effects of suffering that is connected with sexuality can also be seen in how AIDS victims have been treated. In the article "The Happy and the Hopeless" (2013), Jerome Groopman reflects on his time as a doctor working with gay men with AIDS. "From much of society, there was scant sympathy for these suffering men. Fundamentalist preachers thundered that the malady was deserved, a manifestation of God's wrath visited upon sodomites. Many in government expressed no interest in a rare disorder striking what they viewed as a marginal group. And some health care workers shunned the patients, not just out of fear for their own health, but disdain for homosexuals." (Groopman 2013, p. 8)

the survivor begins not only to integrate the traumatic episode into a life with a before and an after, but also to gain control over the occurrence of intrusive memories. When I was hospitalized after my assault I experienced moments of reprieve from vivid and terrifying flashbacks when giving my account of what had happened—to the police, doctors, a psychiatrist, a lawyer, and a prosecutor. Although others apologized for putting me through what seemed to them a retraumatizing ordeal, I responded that it was, even at that early stage, therapeutic to bear witness in the presence of others who heard and believed what I told them. (Brison 2002, pp. 53-54)

4.8 Imagination as a lack of understanding

Throughout this chapter I have questioned the idea that there would be a causal link between a cognitive function or method of imagination, or in other words “empathy”, and our capacity to feel compassion. Even though I do think that imagination is an important aspect of compassion, I have argued that it is problematic to assume that there is a general cognitive function of analogical imagination that enables us to understand other people and to care for them. I have also argued that some of the empirical research on empathy rests on body-mind dualistic assumptions.

In this last part of the chapter I shall discuss one more aspect concerning empathy that is suggested by some theorists. I mentioned in the beginning of this chapter that Martha Nussbaum claims that empathy is a morally

neutral imaginative technique that can be used for *both* good and evil purposes. This is an idea that follows from the assumption that empathy is a cognitive function. It is also an idea that can seem to explain certain forms of cruelty. I shall quote Nussbaum again, but now more at length:

In short, empathy is a mental ability highly relevant to compassion, although it is itself both fallible and morally neutral.

Does empathy contribute anything of ethical importance entirely on its own (when it does not lead to compassion)? I have suggested that it does not: a torturer can use it for hostile and sadistic ends. On the other hand, it does involve a very basic recognition of another world of experience, and to that extent it is not altogether neutral. If I allow my mind to be formed into the shape of your experience, even in a playful way and even without concern for you, I am still in a very basic way acknowledging your reality and humanity [...] Consider Hannibal Lecter's treatment of Clarice Starling in *The Silence of the Lambs*. Although Lecter's intentions toward Clarice are entirely malign, and although he might easily be imagined eating her, nonetheless, in his very effort to reconstruct the workings of her mind there is a basic human respect. The evil of utter dehumanization seems worse: for Jews, or women, or any other victim to be treated as mere objects whose experience doesn't matter may perhaps involve a more profound evil than for them to be tortured by an empathetic villain who recognizes them as human. (Nussbaum 2001, pp. 333-334)

In one sense Nussbaum's reflections on cruelty can seem to be correct. No doubt people can sometimes use their imagination in order to hurt other people. Clearly cruel persons can also sometimes be smart. However, Nussbaum's reflections on empathetic torturers are, I think, still problematic. These problems are on several levels. To begin with, Nussbaum is too uncritical of the film *The Silence of the Lambs* (Tally, 1991). Hannibal Lecter is portrayed as handsome, intelligent and cruel. To portray a murderer as uncommonly intelligent is one way of making cruelty appear exotic. This can be seen in films such as *The Silence of the Lambs*. Nussbaum, however, does not notice that the portrayal of Lecter as highly intelligent and sophisticated is a way of mystifying cruelty and making it appear fascinating and deep in an obscure sense. One way of making sadism exotic is to make out the figure having the lead role in the film, as a hero. Another is to portray him as uncommonly intelligent and handsome, or as uncommonly ugly. My point here is not to say that cruel people are never smart; cruel people can be as smart or as stupid as anyone else. However, we are sometimes seduced to think that cruel persons are "deep" or especially intelligent.

The problem, however, is not merely that Nussbaum is too uncritical of the film, but she takes the film as an example of her ideas of empathy. Nussbaum's reflections on the film are an expression of her cognitive theory about empathy being a neutral imaginative technique that we can use for good or evil purposes. In the quote above Nussbaum distinguishes between a cruel person who

imagines the victim's perspective and a cruel person who treats the victim as an object. She assumes that this is a distinction in a cognitive capacity. At the same time she oddly assimilates the cruel person's capacity to imagine the suffering person's perspective with *respect*, merely because the person imagines "another perspective". I cannot see how her examples of "empathetic torturers" have anything to do with respect. Talking about respect in this sense distorts the meaning of the word.

However, the problem is not only Nussbaum's strange use of the word "respect". My impression is that Nussbaum confuses a moral way of speaking with a cognitive function. Sometimes when people hear about violence they might say such things as "I can't understand". However, when people say so they do not necessarily mean that the cruel person is unusually intelligent or that the cruelty is of a sophisticated character. Nussbaum does not see that when we say that certain cruel acts are impossible to understand we express a moral stance towards what is being done. We are not necessarily describing the cruel person's intelligence. Similarly Nussbaum confuses the moral expression "treating someone like an object" for a cognitive description of how the cruel person is unable to see that other people have minds.

In his essay "Can We Understand Ourselves?" (1997) Peter Winch remarks that he finds it impossible to understand that someone can find football so important

that he is ready to kill for it.⁶² Winch discusses this in connection with questions of how we can understand other cultures. His point here is that our difficulty understanding certain actions may concern things that happen in our own culture, and things that in a way are not complicated. He is not here saying that if he could get more psychological information he might understand such football fanaticism better. Neither is he saying that the football freak is unusually intelligent and that it is therefore difficult to understand his way of reasoning. His saying that he cannot understand is not an expression of a lack of information or a lack of some psychological skill. By saying that it is impossible to understand he is expressing his horror over such a person. Cruelty can have a very simple form and still be impossible to understand. In his paper "The Limits of Understanding" (2005) Hertzberg writes:

⁶² Winch refers here to the murdering of Andrés Escobar. Escobar unfortunately made an own goal during the 1994 FIFA World Cup. It is believed that he was murdered because the goal caused gambling losses among drug lords. It seems that Winch was not aware of the underlying reason for the murder when writing his article. He does not refer to these drug lords but describes the case as if it was a murder committed by football fanatics, which it apparently was not. This can of course be seen as making Winch's point weaker since the murder had a more complex background. Nevertheless, I think his philosophical point is correct and that the example of football suits his point well. Even if Winch was wrong about the real reason for the murder, there is a great deal of fanaticism surrounding football, and because of this violence is not uncommon. It is also a sport that is part of Winch's own culture and thus in no way strange or unfamiliar.

My specific point is this: if I call the activity of the snipers who were shooting the children of Sarajevo incomprehensible, or if I say I do not understand it, this is *not* like saying that I do not like it, or that I find it abhorrent. In fact, I think it would be strange if someone said he did not 'like' the sniping, or that he 'found it abhorrent'. It is as if the distance he is expressing were not deep enough. I should like to say that there is a form of distancing which is the most directly expressed by using words like 'incomprehensible' or 'I don't understand.' It is a reaction to certain forms of evil. (Hertzberg 2005, p. 5)

The sense in which people in wars or in other cases of brutality are treated as if they were not human beings is similar to the sense in which these acts of cruelty may be said to be impossible to understand. These expressions do not necessarily mean that the cruelty is of a very sophisticated psychological character. Nor do these expressions have to do with describing some sort of cognitive-epistemological problems of the cruel person not seeing that the other person has a consciousness.

I claimed earlier that Nussbaum has an exotic conception of cruelty which is connected with her conception of empathy. Exotism is also connected with a distanced conception of violence. C. S. Lewis writes in his *Letters to Malcolm* ([1964] 1991):

[The Crucifixion] did not become a frequent motive of Christian art until the generations which had seen real crucifixions were all dead. As for many hymns and sermons on the subject—endlessly harping on blood, as

if that were all that mattered—they must be the work either of people so far above me that they can't reach me, or else of people with no imagination at all (some might be cut off from me by both these gulfs). (Lewis [1964] 1991, p. 86)

According to Lewis, one can see a difference between how persons who witnessed real crucifixions portrayed it and how it has been portrayed later on when crucifixions have become a historical curiosity. He suggests that a certain kind of detailed imagination of things such as torture is not necessarily an expression of insightful understanding or an expression of a true ability to imagine what suffering means. This kind of detailed imagination about certain forms of suffering may rather reflect a *lack* of imagination. It can be reflective of a lack of imagination in the sense that it is a lack of ability to be truly moved by human life; a curiosity about abhorring details.⁶³

One might also say that Lewis' here distinguishes between two forms of what we mean when we say that a person has imagination. On the one hand a person can have a *detailed* imagination. On the other hand we can also sometimes say that such a person *lacks* imagination. The former way of talking about imagination could be described as a cognitive capacity. The latter could again be described as a moral stance. However, Lewis does not merely distinguish between these two meanings of how we talk of imagination. He suggests that the cognitive capacity

⁶³ An example of such lack of imagination would for instance be Fitzgerald et al.'s film *The Passion of the Christ* (2004).

is not neutral in kind but can in itself be expressive of our attitude to what is being done. Our ability to think shows itself in how we react to cruelty: in our finding it difficult to talk about, but also in our trying to acknowledge cruelty by talking about it, as well as in our trying to help others. And sometimes being able to imagine and talk about all the gruesome details of torture with ease can be seen as an expression of a superficial attitude towards life, a lack of imagination. Even if the detailed imagination of suffering is not an expression of contempt or hate or any other clearly destructive attitude, the interest in picturing the suffering as exactly and vividly as possible may be an expression of blindness. Susan Sontag writes in her book *Regarding the Pain of Others* (2003):

It seems that the appetite for pictures showing bodies in pain is as keen, almost, as the desire for ones that show bodies naked. For many centuries, in Christian art, depictions of hell offered both of these elemental satisfactions. (Sontag 2003, p. 41)

Sontag further notes that images of violence and suffering tend to have an especially shocking character when it comes to reporting about war or famine in Asia or Africa.

Generally, the grievously injured bodies shown in published photographs are from Asia or Africa. This journalistic custom inherits the centuries-old practice of exhibiting exotic—that is colonized—human beings. (Sontag 2003, p. 72)

These images of suffering people in Asia and Africa are generally combined with compassionate descriptions. However, these compassionate descriptions have an ambiguous character. Sontag further argues that the frequent portrayals of conflicts in Africa and Asia do not mean that these conflicts actually get the attention they ought to get. Rather the abundance of images of violated people may feed an attitude of passivity, a feeling that this is simply how it is there and how it will always be in those parts of the world. We might feel compassion for the people suffering in these countries, but it can be an empty passive form of compassion, a compassion that does not bother us in our sleep. This attitude of passive compassion is not very far from curiosity. It is compassion as a form of entertainment.

In *Elizabeth Costello* (2003), J.M. Coetzee also reflects on the relation between imagination and cruelty. The book tells about an aged writer named Elizabeth Costello. Costello reflects, among other things, on how we can become seduced by literature which is seeped in detailed descriptions of torture. Costello's suggestion is that simply reading such things will, in a sense, make us part of this evil. I think Coetzee in this novel says something important about imagination. He makes us think about imagination not merely as a way of reflecting which does not touch on who we are, and which does not have any moral relevance. The way we imagine things, and also the way we can be seduced into imagining things, expresses and shapes who we are. This does not mean that it would be wrong to reflect on cruelties in war or other forms of

cruelty. That would be an absurd consequence. But, as Sontag notes, the difficulty lies in being able to describe such atrocities without being seduced by them. Sontag also reflects on her earlier work, *On Photography* (1977) where she discusses how news reports on sufferings in war have become a form of entertainment.

The question [in *On Photography*] turns on a view of the principal medium of the news, television. An image is drained of its force by the way it is used, where and how often it is seen. Images shown on television are by definition images of which, sooner or later, one tires. What looks like callousness has its origin in the instability of attention that television is organized to arouse and to satiate by its surfeit of images. [...] The whole point of television is that one can switch channels, that it is normal to switch channels, to become restless, bored. Consumers droop. They need to be stimulated, jump-started, again and again. (Sontag 2003, pp. 105-106)

Sontag argues that the news reports on television media are designed so as to appeal to a kind of light attention; a form of attention that can be turned on or turned off or changed whenever one feels like it. Here knowledge becomes something we can choose freely to reflect on or to forget. In this sense it becomes a kind of entertainment.

Sontag also notes that whether a conflict becomes something we think about and remember depends on the larger political impact of the conflict as well as on its proximity. She names the Spanish civil war, the Serb and

Croat wars in Bosnia and the Israeli - Palestinian conflict as examples of wars that have got a lot of attention. However, she continues:

In the meantime, far crueller wars in which civilians are relentlessly slaughtered from the air and massacred on the ground (the decades-long civil war in Sudan, the Iraqi campaigns against the Kurds, the Russian invasions and occupation of Chechnya) have gone relatively underphotographed. (Sontag 2003, pp. 36-37)

Often war and devastation go on for many years and slowly we tend to lose our interest in them. Slowly the news reports also fade away because there are no “news”, everything goes on as usual, villages are bombed and people are killed and slaughtered as usual. Conflicts fall out of our daily reflections and worries even though we know they continue. This is of course only a reaction to suffering that is possible if you are at a safe distance.

However, Sontag is also critical of a tendency today to claim that all reality has become mere entertainment because of media.

According to a highly influential analysis, we live in a ‘society of spectacle.’ Each situation has to be turned into a spectacle to be real—that is, interesting—to us. [...] There are only representations: media. (Sontag 2003, p. 109)

She further comments:

To speak of reality becoming a spectacle is a breathtaking provincialism. It universalizes the viewing habits of a small, educated population living in the rich part of the world, where news has been converted into entertainment [...] It assumes that everyone is a spectator. It suggests, perversely, unseriously, that there is no real suffering in the world. (Sontag 2003, p. 110)

As Sontag points out, the thought that media per se distort all of reality into a mere spectacle is expressive of cynicism. Cynicism is also one form of attitude towards knowledge, an attitude closely connected with passivity.

4.9 Conclusion

I began the chapter by discussing some psychological experiments on the relation between empathic imagination and compassion. The experimental research on empathy is taken to prove that there is a causal link between the two. I questioned the experiments on basically two accounts. First, the experimental context is so strongly reduced that it is unclear in what sense the participants can be said really to understand that another person is suffering. Thereby it also becomes unclear how one should understand their compassionate responses. Second, the instructions that the experimenters give to the test subjects are leading. Thereby it also becomes unclear how the responses should be understood.

However, the experiments are not merely problematic because they lack a meaningful context or because the

questions are leading. The character of the experiments is also reflective of a conceptual confusion. The experiments reflect the assumption that compassion depends on the capacity to use our imagination in an analogical sense. This idea is expressive of a tendency to consider the second person perspective as dependent on a first-person perspective, when we talk of understanding. A tendency to describe interpersonal understanding in a body-mind dualistic manner is also evident in the experiments.

Another assumption that shapes the theories on the relation between empathy and compassion is the idea that other people have separate perspectives in some general cognitive sense. The assumption then is that we need to use our imagination in order to bridge this mental gap. In contrast to this, Emmanuel Levinas and Lars Hertzberg argue that the question of separateness ought to be understood in a moral sense, not in a cognitive sense. I have a *responsibility* to care for the suffering person. As Hertzberg argues, the way we can reflect and worry about a suffering person, and the way we can feel that we do not know what he experiences, is in itself expressive of our sense for the seriousness of his situation. The difference in perspective is not a general cognitive and epistemological dilemma that can be solved by a method of imagination.

By discussing several biographical descriptions of suffering as well as of compassion I have tried to bring forth how experiences of suffering affect interpersonal relationships in many ways. The descriptions suggest that our difficulties in understanding another person's suffering, including our sense of separateness and

estrangement, are of a quite different (relational and moral) character than the general cognitive obstacles that are taken for granted in the psychological and philosophical theories on empathy.

In the last part of the chapter I discussed the idea that empathic imagination can be used for both good and evil purposes. The assumption that we can do so rests on a conception of interpersonal understanding as a neutral cognitive device. I argued that such a conception of understanding is problematic. Knowledge, or vivid imagination, is not necessarily always expressive of understanding. Our ways of knowing all the details of certain forms of suffering can sometimes be an expression of shallow curiosity. There is, so to say, no general neutral way of understanding other people's suffering. The ways we think about suffering, the ways we understand things, are expressive of our attitude towards the suffering person. In this sense our understanding of suffering cannot be separated from the fact that we are responsible for each other's life.

Concluding remarks

A main theme of this thesis has been to look at the relation between theory-of-mind theories and empirical research. Theories on theory of mind often lean heavily on empirical research and observations within psychology, psychiatry, as well as within other scientific fields. There are also many observations in real life that can seem to support theory-of-mind theories. This connection with empirical research and real life observations makes theory-of-mind theory influential within both philosophy and psychology. However, I have argued that this relation between theory and empirical observations is not as clear as it can seem. Much of the empirical research is built on certain theoretical assumptions that shape the research and that also shape the results. I have argued that theory-of-mind theories rest on the assumption that interpersonal understanding is a cognitive matter. The cognitive conception of understanding is further connected with the assumption that interpersonal understanding basically is epistemological in kind. That is, it is assumed that to understand another person means to have knowledge about the person. These assumptions are reflected in several more specific (problematic) forms of explaining interpersonal understanding.

First, there is the tendency to consider a third-person perspective as basic for what it means to understand other people. This can be discerned in how the researchers tend to talk of interpersonal understanding as, for instance, a

matter of predicting, explaining or attributing mental states to other people. That is, a distanced, observing perspective, is assumed to be basic for interpersonal understanding. I have also argued that the third-person perspective forms how psychological empirical research that is connected with theory-of-mind theory is built up and how results are interpreted.

Second, I have argued that there is a tendency to consider interpersonal understanding as based on an analogical mechanism of imagination or as an analogical bodily reaction. This is a central assumption in theories concerning altruism, infant imitation, mirror neurons and empathy. The idea that interpersonal understanding is based on an analogical mechanism of imagination (or based on analogical bodily reactions), is problematic because it is assumed that we can only understand other people indirectly, while it is assumed that we have privileged access to our own feelings. I also claim that these assumptions form the empirical research.

Further, there is a tendency to consider the human being in body-mind dualistic terms. Since interpersonal-understanding is assumed to consist in a third-person perspective, also the human body is described from a third-person perspective, as something we *observe*. A body-mind dualistic perspective can also be discerned in connection with the assumption that interpersonal understanding consists in analogical imagination. If one assumes that interpersonal understanding takes an analogical form, it becomes important that we have physically similar bodies. However, body-mind dualism

also reflects the fact that psychology has a history of influence from physiology.

I claim that these three above described features; a third-person perspective, the idea that interpersonal understanding consists in analogical imagination and body-mind dualistic perspectives, are central in theory-of-mind theories.

In the first chapter, on altruism, I contrasted a rationalistic perspective with emotivistic perspectives on altruism. According to the rationalistic perspective, which is represented by Leda Cosmides and John Tooby, human social life is based on a nonconscious algorithmic reasoning mechanism by which we calculate how to act in order to maximize fitness when dealing with others. According to the emotivistic perspective, on the other hand, which is represented by Adam Smith, Charles Darwin, Edward Westermarck and Frans de Waal, human social life is based on a capacity for analogical imagination and an analogical emotional responsiveness. Both perspectives lean on convincing empirical observations about human and animal behaviour. Still, I argue that both theoretical perspectives are based on philosophical confusions. These problems can be summed up as reliance on a transactional model of interpersonal relationships on the one hand, and the view of interpersonal understanding as consisting in an ability for analogical imagination on the other hand. Even though human reasoning often has a transactional, calculating character I have argued that it is problematic to assume that human social life and interpersonal understanding are in some general sense based on a

transactional manner of reasoning. Further, even though human beings often respond to other people in a kind of analogical emotional manner, and even though our ways of imagining can take an analogical form, it is problematic to assume that interpersonal understanding in a general sense is based on an analogical mechanism of emotional responsiveness. The argument from analogy also rests on conceptual confusion. It is assumed that the second-person perspective is dependent on the first-person perspective.

Further, I have claimed that even though the rationalistic and the emotivistic perspectives can seem to be opposites, they basically share certain cognitive and also epistemological conceptions of interpersonal understanding. Cosmides and Tooby's perspective on human social life as based on a nonconscious transactional reasoning mechanism, does not necessarily contradict the theories involving empathy.

Both perspectives also share the assumption that human social life originates in a need to cooperate. Clearly cooperation (and transaction) is a central aspect of human life. However, I have argued that the idea that human social life originates in a need for cooperation (or transaction) does not show what it means for human beings to stand in personal relationships, relationships that cannot merely be defined as having an instrumental meaning. I have also claimed that it is problematic to suggest that our emotional responsiveness to each other has evolved because it enhances cooperation. The natural aspects of human social life, such as our caring for our children and our also caring for our loved ones when they

are dying, cannot be understood if one does not acknowledge the ways in which our long-term close personal relationships, and accordingly also language, give meaning to our actions. In this sense the concepts of cooperation or transaction that the rationalistic and emotivistic theorists work with rest on a reductive image of the natural character of human social life.

In chapter two I dealt with certain psychological experiments concerning imitation that have been made with young children. Psychologists have shown that newborn infants have a capacity to imitate other people's facial expressions. These findings have been taken to imply that children have an innate theory-of-mind function by which they learn to decipher other people's facial expressions and thereby also eventually to understand other people's intentions. One of my aims in this chapter was to discuss and question how these empirical tests concerning imitation have been constructed. My aim was not to deny the empirical findings *per se*. It seems clear that infants do occasionally tend to imitate other people's facial expressions. Rather I have claimed that the researchers give this feature a too dominating importance when it comes to interpersonal understanding. Since the imitation test is so highly restricted in its focus it creates the impression that imitation is a specifically important response that eventually enables the child to learn to understand others. This creates, again, the impression that there is some such thing as a theory-of-mind function. I have claimed that it is problematic to regard the infant's occasional imitation as the *only* or the *main* factor that leads

to interpersonal understanding. In contrast I have argued that it is important to regard the child's responsiveness (including occasional tendencies to imitate facial expressions) as part of a broader dialogical relationship between parent and child.

I have also argued that the experiments concerning imitation are constructed in a one-directional manner. The adult tries to elicit certain imitative responses in the infant, but the adult is not *responding* to the infant. Nor is it clear that the adult is trying to engage with the child in a meaningful way. Bronfenbrenner maintains that the strong influence of natural scientific research methods on psychology have formed experimental research practice so that it takes a one-directional form. However, I also argue that the one-directional character of the psychological experiment fits well with an epistemological conception of understanding which theory-of-mind theory is based on. It is because the researchers have an epistemological and also analogical conception of interpersonal understanding that they give so much importance to infants' capacity to imitate facial expressions. I further argued that a similar inclination to construct empirical studies on interpersonal understanding in a one-directional manner can be seen in Paul Ekman's studies on facial expressions as well as in research on mirror neurons.

Certain themes come up both in chapters one and two. In both chapters I claim that there is a close, and problematic, relation between certain empirical observations and philosophical theory. In chapter one I contrasted a cognitive rationalistic perspective on altruism

with a cognitive emotivistic perspective on altruism as consisting of analogical imagination. Both these also rest on an epistemological conception of understanding. I further claimed that both the rationalistic and the emotivistic theories rest on an instrumental conception of human relationships. Also in chapter two I addressed the idea that interpersonal understanding is based on an analogical responsiveness.

In chapter three, on autism and theory of mind, I discussed the claim that autism is due to a so called theory-of-mind deficit or, in other words, "mindblindness". The fact that persons with autism generally have large difficulties in understanding other people and the fact that they often seem to be unaware of other people, seems to imply that they lack a theory-of-mind function. Researchers who support the theory that persons with autism lack a theory-of-mind function have focused on certain specific aspects concerning autism. First, it is claimed that children with autism are aware of objects but unaware of people. Second, since children with autism are not good at pretend play it is claimed that this points to an inability to use an analogical method of imagination, which is reflective of a lack of a theory-of-mind function. Third, since children with autism fail in the so called "false belief task" this is assumed to indicate that they cannot see that other people have intentions, or in other words, that they lack a theory-of-mind function. Fourth, it is claimed that the fact that people with autism can have a tendency for echolalia when talking also indicates that they lack a theory-of-mind function. Fifth, it is claimed that since

children with autism are not good at understanding humor and irony it implies that they lack a theory-of-mind function. In this sense there is much empirical evidence that seems to point to the absence of a theory-of-mind function.

In the chapter I discuss the above mentioned aspects of autism that theory-of-mind theorists take to imply that autism can be explained as mindblindness. Even though the features described above are often present in autism, it is problematic to take these as evidence for mindblindness or for the absence of a theory-of-mind function. Furthermore I argue that there is a tendency, among proponents of the theory-of-mind view, to focus merely on certain forms of behaviour, while ignoring other forms. By discussing Clara Claiborne Park's biographical books (1967) (2001), I have tried to present a more varied picture of autism.

In this chapter I also, once again, addressed the tendency to consider interpersonal understanding as an epistemological tool. Much of the empirical observations, including certain psychological tests, are shaped by the assumption that our understanding of other people is a matter of learning to predict and explain behaviour. In this sense an epistemological third-person perspective largely shapes the empirical observations concerning autism. I have argued that this is a problematic assumption concerning what it means to understand other people. Further, the emphasis on pretend play reflects the assumption that we learn to understand other minds by

the use of analogical imagination. This idea comes up for discussion in all of the chapters in some way or other.

However, the problem is not merely that the theory-of-mind theorists give a one-dimensional picture of autism but also that they give a problematically intellectualistic picture of normal understanding, something that, among other things, is reflected in the conception of irony that the theorists work with. In this sense the aim of this chapter is twofold. One aim has been to discuss and question the claim that the problems in understanding that persons with autism struggle with, can be defined as a lack of a theory-of-mind function or that they are “mindblind”. By this I have *not* meant to say that persons with autism have no problems in understanding. The other aim has been to question the assumption that normal forms of interpersonal understanding could be defined as consisting of a well-functioning theory of mind.

In the fourth, and last, chapter I have discussed certain theories claiming that there is a relationship between empathy and compassion. I began the chapter by discussing and questioning some psychological experiments on the relation between empathic imagination and compassion. The experimental research on empathy is taken to prove that there is a causal link between the two. I have questioned the experiments on basically two accounts. First, the experimental context is so strongly reduced that it is unclear in what sense the participants can be said really to understand that another person is suffering. Thereby it also becomes unclear how one should understand their compassionate responses. Second, the instructions that the

experimenters give the test subjects are leading. Thereby it also becomes unclear how the responses should be understood.

However, the experiments are not problematic merely because they lack a meaningful context or because the questions are leading. The experiments also reflect a conceptual confusion. They reflect the assumption that compassion depends on the capacity to use our imagination in an analogical sense. This assumption is connected with the assumption that the second-person perspective is dependent on the first-person perspective. This makes the researchers think that I can only understand another person's feelings if I imagine that I have that feeling myself.

In this sense certain problematic patterns of explaining interpersonal understanding are apparent in these theories; patterns that I have also argued can be discerned in the other chapters. Further, a tendency to describe interpersonal understanding in a body-mind dualistic manner can be seen in the experiments. Body-mind dualistic explanations of interpersonal understanding have also come up for discussion in earlier chapters.

Another assumption that shapes the theories on the relation between empathy and compassion is the idea that other people have different perspectives in some general cognitive sense. The suggestion then is that we need to use our imagination in order to bridge this cognitive gap. In contrast to this, Emmanuel Levinas and Lars Hertzberg argue that the question of difference in perspective ought to be understood in a moral sense, not in a cognitive sense.

I have a *responsibility* to care for the suffering person. As Hertzberg argues, the way we can reflect and worry about a suffering person, and the way we can feel that we do not know what he experiences, is in itself expressive of our sense of the seriousness of his situation. It is not a general cognitive dilemma that can be solved by a method of imagination.

By discussing several biographical descriptions of suffering as well as of compassion I have tried to bring forth how experiences of suffering affect interpersonal relationships in many ways. The descriptions suggest that our difficulties in understanding another person's suffering, including our sense of separateness and estrangement, are of a quite different (relational and moral) character than the general cognitive epistemological obstacle that is taken for granted in the psychological and philosophical theories on empathy.

In the last part of the chapter I have discussed the idea that empathic imagination can be used for both good and evil purposes. This idea rests on a conception of interpersonal understanding as a neutral epistemological device. I have argued that such a conception of understanding is problematic. Knowledge, or vivid imagination, is not necessarily always expressive of understanding. There is no general neutral way of understanding other people's suffering. The ways we think about suffering, the ways we understand things, are expressive of our attitude towards the suffering person.

To conclude then, I have in this thesis discussed quite a variety of topics ranging from evolutionary theories of

altruism, to theories concerning infant imitation, to theories about autism, to theories of empathy. I have argued that there are certain problematic patterns of theoretical thinking that tend to recur in the varying fields of research. I have also claimed that theory-of-mind theories are influential in many areas of scientific discussion. Theory-of-mind theories influence empirical research. I have, however, argued that the empirical research is often more ambiguous than might appear at first sight. A further reason why theory-of-mind theories are influential is because the theories reflect certain classical philosophical ideas. The empirical research is often formed in a one-directional and reductive way that fits well with a cognitive and epistemological conception of interpersonal understanding. Epistemological conceptions of interpersonal understanding can take many forms such as an emphasis on a third-person perspective, the argument from analogy and body-mind dualism; ideas which all have a long history within philosophy. Further, the empirical research as well as the theoretical assumptions are supported by many real life observations.

A central aim of the thesis has been to discuss and question the impression that theory-of-mind theories describe a basic underlying cognitive function of interpersonal understanding. I have questioned this impression by pointing at certain conceptual confusions that shape the theories as well as the empirical research. The impression that theory-of-mind theories describe basic underlying cognitive functions of understanding is created by a recurring tendency towards certain kinds of

conceptual confusions, combined with carefully made and convincing empirical research that, despite this, is formed by these conceptual confusions, combined with the fact that certain real life descriptions of human life seem to fit well with the theories.

Swedish summary - Sammanfattning

En populär idé inom dagens filosofiska och psykologiska forskning om interpersonlig förståelse, är idén att vi använder en kognitiv funktion (eller metod) för att förstå andra människor, en så kallad "theory of mind" funktion. Denna idé förekommer inom ett brett vetenskapligt fält så som inom evolutionspsykologi, inom teorier om barns utveckling, inom teorier om autism, samt inom emotionsfilosofi och moralfilosofi.

Avsikten i denna studie är att se närmare på vissa inflytelserika filosofiska och psykologiska teorier om interpersonlig förståelse, som också har en stark koppling till empirisk forskning. I arbetet hävdar jag att teorierna ifråga avspeglar vissa klassiska, problematiska filosofiska antaganden. Dessa antaganden bestämmer hur teorierna byggs upp, de formar hur exempel beskrivs samt också hur empiriska undersökningar byggs upp och hur resultat tolkas.

Ett genomgripande antagande i teorierna ifråga är att interpersonlig förståelse i grundläggande mening är av epistemologisk karaktär. Det vill säga, man tänker sig att vår förståelse av andra (i grunden) handlar om att försöka få *information* om den andras tankar och känslor. Detta epistemologiska perspektiv på interpersonlig förståelse bygger på ytterligare tre klassiska filosofiskt problematiska tankegångar. 1. Forskningen utgår från att ett tredjepersons perspektiv är grundläggande för vår förståelse av andra människor. D.v.s. man utgår från att då

vi förstår andra människor så *betraktar* vi dem som neutrala observatörer på avstånd och *tänker på* dem. 2. Forskningen utgår från att vår förståelse av andra människor bygger på en generell kognitiv förmåga till analogiskt tänkande, eller en förmåga att reagera kroppsligen på ett analogt sätt. Denna idé präglar både teorier om altruism, teorier om spädbarns förmåga att imitera ansiktsuttryck, teorier om så kallade spegelneuroner, samt teorier om empati. 3. Det finns en återkommande tendens i forskningen ifråga att beskriva människan i dualistiska, kropp/själ, termer. Trots en omfattande, och övertygande, empirisk forskning, hävdar jag att denna ofta på ett missledande sätt bestäms av de ovan nämnda perspektiven.

Avhandlingen består av en längre introduktion samt fyra kapitel. I det första kapitlet "Rationalistic and Emotivistic Theories of Altruism", kontrasterar jag två evolutionsteoretiska perspektiv på altruism. Jag inleder med att diskutera psykologen Leda Cosmides och antropologen John Toobys idé att altruism grundar sig på en algoritmisk tankefunktion. Med hjälp av denna tankefunktion antas vi omedvetet kalkylera hur vi bör handla för att maximera vår egen vinning då vi umgås med andra människor. I kontrast till denna idé lyfter jag fram primatologen Frans de Waals evolutionsteoretiska teori om altruism som något som bygger på en emotionell empatisk förståelseförmåga. Jag hävdar att Cosmides och Toobys liksom de Waals perspektiv bygger på två slags problematiska antaganden. För det första bygger de på antagandet att interpersonliga relationer i grundläggande mening kan beskrivas som transaktionella

relationer. För det andra bygger de på antagandet att vår förståelse av andra människor i en generell mening grundar sig på en kognitiv förmåga att föreställa oss själva i den andras situation (det så kallade analogiargumentet).

I det andra kapitlet, "Theory of Mind and Infant's Imitation of Facial Expressions", diskuterar jag ett perspektiv på den mänskliga kroppen som finns inom så kallad simulationsteori. Enligt simulationsteori bygger vår förståelse av andra människor på en analogisk föreställningsförmåga, vilken å sin sida är beroende av en förmåga att simulera andra människors kroppsliga beteende. I kapitlet diskuterar jag vissa psykologiska experiment där man studerat spädbarns förmåga att imitera andra människors ansiktsuttryck. Experimenten påvisar att spädbarn har en förmåga att imitera andra människors ansiktsuttryck. Enligt psykologen Andrew N. Meltzoff visar dessa resultat att spädbarn har en kognitiv analogisk föreställningsförmåga att förstå andra. Med hjälp av denna kognitiva funktion lär sig barn att förstå andra människors ansiktsuttryck och så småningom då också att förstå andra människors avsikter. I kapitlet diskuterar jag hur dessa experiment skall förstås. Min avsikt är inte att bestrida de empiriska resultaten i sig. Däremot hävdar jag att teorin och också experimenten reflekterar vissa klassiska, problematiska, filosofiska antaganden om interpersonlig förståelse. Dessa antaganden styr hur experimenten byggts upp samt hur resultaten tolkas.

I kapitel tre, "Autism and Theory of Mind", diskuterar jag tanken att autismspektrumtillstånd beror på avsaknad av en så kallad "theory of mind" funktion.

Autismspektrumtillstånd (härefter autism) är en genomgripande utvecklingsstörning som påverkar en människas liv på många vis. Det mest slående draget i autism är avsaknaden av social responsivitet. I kapitlet diskuterar jag antagandet att autism beror på en avsaknad av en "theory of mind" funktion. Jag betraktar närmare ett flertal empiriska studier som verkar stöda denna teori. I kontrast till de empiriska studierna diskuterar jag också Clara Claiborne Parks (1967), (2001) biografier om sin autistiska dotter. Även om de empiriska studierna pekar på olika former av social dysfunktion, hävdar jag att resultaten av de empiriska undersökningarna ifråga är mer mångtydiga än vad forskarna uppmärksammar. Min avsikt är inte att förneka att barn med autism har olika slags svårigheter att förstå andra människor. Jag hävdar dock att dessa svårigheter inte kan förstås från synvinkeln att människors förståelse av varandra skulle bygga på en "theory of mind" funktion.

I kapitel fyra, "The Relationship Between Empathic Imagination and Compassion", diskuterar jag idén att det finns en kausal relation mellan en så kallad empatisk föreställningsförmåga och medlidande (vår vilja att hjälpa). Jag inleder med att diskutera vissa psykologiska experiment som anses visa att det finns en kausal koppling mellan en empatisk förmåga att föreställa sig själv i den andras situation och medlidande. Å ena sidan hävdar jag att de experimentella situationerna är så starkt reducerade att det blir oklart på vilket vis försökspersonernas reaktioner kan sägas vara uttryck för medkänsla (eller vara uttryck för avsaknad av medkänsla). Vidare hävdar jag att

instruktionerna som ges åt försökspersonerna är ledande. Detta innebär att resultaten blir ytterligare mångtydiga.

Genom att lyfta fram ett antal biografiska beskrivningar av lidande hävdar jag att våra svårigheter att förstå andra människors lidande är av en annan karaktär än vad forskarna tar för givet i experimenten och i teorierna om empati. Samtidigt finns det dock vissa drag i vårt liv och i hur vi kan reagera på andra människors lidande, som delvis kan förklara varför de kognitiva teorierna om empati är lockande.

Min avsikt i avhandlingen är således att undersöka "theory of mind" teorier och deras relation till empirisk forskning samt också relationen mellan dessa filosofiska teorier och iakttagelser av vardagslivet. Jag hävdar att vissa klassiska filosofiska antaganden ofta reflekteras både i den empiriska forskningen liksom även i iakttagelserna av vårt vardagsliv, utan att forskarna ifråga är medvetna om detta. Även om jag undersöker, och också ifrågasätter, den empiriska forskningen som berör forskning kring "theory of mind", är min avsikt inte att bidra med ny empirisk forskning. Mitt syfte är att påvisa återkommande begreppsliga förvirringar vilka bestämmer teorierna ifråga samt vilka också påverkar hur den empiriska forskningen byggs upp och hur resultaten tolkas.

Bibliography

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*. Arlington, VA: American Psychiatric Association.

Baron-Cohen, S. T. (1997). *Mindblindness: An Essay on Autism and Theory of Mind*. London: The MIT Press.

Baron Cohen, S., Leslie, A. M. and Frith, U. (1985). "Does the Autistic Child have a 'Theory of Mind'?" *Cognition*, 21: 37-46.

Baron-Cohen, S. T. and Swettenham, J. (1997). "Theory of Mind in Autism: Its Relationship to Executive Function and Central Coherence", in *Handbook of Autism and Pervasive Developmental Disorders*. Second edition. Edited by D. Cohen and F. Volkmar. Hoboken: John Wiley & Sons.

Batson, D. C. (1991). *The Altruism Question, Towards a Social-Psychological Answer*. Hillsdale: Lawrence Erlbaum Associates.

Batson, D. C., Sager, K., Gast, E., Kang, M., Rubchinsky, K., and Dawson, K. (1997). "Is Empathy-Induced Helping Due to Self-Other Merging?" *Journal of Personality and Social Psychology*, 73: 495-509.

Bernard-Opitz, V. (1982). "Pragmatic Analysis of the Communicative Behavior of an Autistic Child" *Journal of Speech and Hearing Disorders*, 47: 99-109.

Biklen, D. (2005). *Autism and the Myth of the Person Alone*. New York: New York University Press.

Bloom, P. and German, T. P. (2000). "Two Reasons to Abandon the False Belief Task as a Test of Theory of Mind" *Cognition*, 77: B25-B31.

Bowlby, J. ([1969] 1997). *Attachment and Loss. Volume 1: Attachment*. London: Pimlico.

Brison, S. (2002). *Aftermath, Violence and the Remaking of a Self*. Princeton: Princeton University Press.

Bronfenbrenner, U. (1978). *The Ecology of Human Development, Experiments by Nature and Design*. Cambridge: Harvard University Press.

Buytendijk, F. J. J. (1962). *Pain: Its Modes and Functions*. Translated by E. O'Shiel. Chicago: University of Chicago Press.

Cavell, S. ([1969] 2002). "Knowing and Acknowledging" in *Must we Mean What we Say?*. New York: Cambridge University Press.

Cockburn, D. (1985). "The Mind, the Brain and the Face" *Philosophy*, 60: 478-493.

Cockburn, D. (1994). "Human Beings and Giant Squids" *Philosophy*, 69: 135-150.

Cockburn, D. (2009). "Emotion, Expression and Conversation" in *Emotion and Understanding: Wittgensteinian Perspectives*. Edited by Y. Gustafsson, C. Kronqvist and M. McEachrane. Basingstoke: Palgrave Macmillan.

Coetzee, J. M. (2003). *Elizabeth Costello*. London: Penguin Books Ltd.

Colliander, T. (1956). *Samtal med smärtan* [Conversations with Pain]. Helsingfors: Söderströms.

Comte, A. (1858). *The Catechism of Positive Religion*. Translated by R. Congreve. London: Chapman.

Cosmides, L. and Tooby, J. (1992). "Cognitive Adaptations for Social Exchange" in *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, edited by J. H. Barkow, L. Cosmides, and J. Tooby. New York: Oxford University Press.

Cosmides, L. and Tooby, J. (2008). "Can a General Deontic Logic Capture the Facts of Human Moral Reasoning? How the Mind Interprets Social Exchange Rules and Detects

Cheaters" in *Moral Psychology, Vol 1, The Evolution of Morality: Adaptations and Innateness*, edited by W. Sinnott-Armstrong. Cambridge: The MIT Press.

Currie, G. (1995). "Imagination and Simulation: Aesthetics Meets Cognitive Science" in *Mental Simulation*, edited by M. Davies and T. Stone. Oxford: Blackwell Publishers.

Darwin, C. ([1859,1871] 1872). *The Origin of Species and The Descent of Man*. Sixth edition. New York: The Modern Library.

Darwin, C. (1872). *The Expression of Emotions in Man and Animals*. London: John Murray.

Dennett, D. (1978). "Beliefs about Beliefs", *Behavior and Brain Sciences*, 4: 568-570.

De Waal, F. (2006). "Morally Evolved, Primate Social Instincts, Human Morality and the Rise and Fall of 'Veneer theory'" in *Primates and Philosophers, How Morality Evolved*, edited by S. Macedo and J. Ober. Princeton: Princeton University Press.

De Waal, F. (2009). *The Age of Empathy, Nature's Lessons for a Kinder Society*. New York: Harmony Books.

Dilman, I. (1987). *Love and Human Separateness*. Oxford: Basil Blackwell Ltd.

Donald, M. (1991). *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Harvard: Harvard University Press.

Ebersole, F. (2001). *Things we Know*. Second edition. Bloomington: Xlibris.

Ekman, P. (1993). "Facial Expression and Emotion" *American Psychologist*, 48, 4, April: 384-392.

Ekman, P. and Wallace, V. F. (1971). "Constants Across Cultures in the Face and Emotion" *Journal of Personality and Social Psychology*, 17, 2: 124-129.

Fay, W. H. (1973). "On the Echolalia of the Blind and of the Autistic Child." *Journal of Speech and Hearing Disorders*, 38: 478-489.

Flack, J. C. and de Waal, F. (2000). "'Any Animal Whatever' Darwinian Building Blocks of Morality in Monkeys and Apes" in *Evolutionary Origins of Morality: Cross Disciplinary Perspectives*, edited by L.D. Katz. Thorverton: Imprint Academic.

Fiser, K. (1986). "Privacy and Pain" *Philosophical Investigations*. 9, 1: 1-17.

Fitzgerald, B., Gibson, M. and Fulco, W. (2004). *The Passion of the Christ*. Directed by M. Gibson. USA: Newmarket films.

Freud, S. (1911). "Formulation on the two principles of mental functioning", in *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Vol. 12*, edited by J. Strachey. London: Hogarth Press.

Gallagher, H. L. and Frith, C. D. (2003). "Functional Imaging of 'Theory of Mind'" *Trends in Cognitive Sciences*, 7, 2, February: 77-83.

Gallese, V., Keysers, C. and Rizzolatti, G. (2004), "A Unifying View of the Basis of Social Cognition" *Trends in Cognitive Science*, 8: 396-403.

Goldie, P. (2000). *The Emotions, a Philosophical Exploration*. New York: Oxford University Press.

Goldie, P. (2006). "Wollheim on Emotion and Imagination" *Philosophical Studies*, 127:1-17.

Goldman A. I. (1995). "Empathy, Mind and Morals" in *Mental Simulation*, edited by M. Davies and T. Stone. Oxford: Blackwell Publishers Ltd.

Goldman A. I. (2006). *Simulating Minds: The Philosophy, Psychology and Neuroscience of Mindreading*. Oxford: Oxford University Press.

Gopnik, A. (1993). "Mindblindness". Unpublished essay. University of California, Berkeley.

Groopman, J. (2013). "The Happy and the Hopeless" *The New York Review of Books*, February 7, 60, 2. <http://www.nybooks.com/articles/archives/2013/feb/07/happy-hopeless/?insrc=toc>

Gupta, A. and Ferguson, J. (1992). "Beyond 'Culture': Space, Identity and the Politics of Difference" *Cultural Anthropology*, 7, 1: 6-23.

Gustafsson, Y. (2009). "Illusions of Empathy", in *Emotions and Understanding: Wittgensteinian Perspectives*. Edited by Y. Gustafsson, C. Kronqvist and M. McEachrane. Houndmills: Palgrave Macmillan.

Gustafsson, Y. (2012). "Three Perspectives on Altruism", in *Language, Ethics and Animal Life: Wittgenstein and Beyond*. Edited by N. Forsberg, M. Burley and N. Hämäläinen. New York: Bloomsbury.

Hamilton, W. D. (1964). "The Genetical Evolution of Social Behaviour II." *Journal of Theoretical Biology*, 7: 17-52.

Hamlyn, D. W. ([1983] 1994). *Perception, Learning and the Self, Essays in the Philosophy of Psychology*. Hampshire: Gregg Revivals.

Hamlyn, D.W. (1978). *Experience and the Growth of Understanding*. London: Routledge & Keegan Paul Ltd.

Happé, F. (1991). "The Autobiographical Writings of Three Asperger Syndrome Adults: Problems of Interpretation and Implications for Theory", in *Autism and Asperger Syndrome*, edited by U. Frith. Cambridge: Cambridge University Press.

Happé, F. (1993). "Communicative Competence and Theory of Mind in Autism: A Test of Relevance Theory" *Cognition*, 48: 101-119.

Heal, J. (1995). "How to Think about Thinking", in *Mental Simulation*, edited by M. Davies and T. Stone. Oxford: Blackwell Publishers.

Hertzberg, L. (2002). "Om inlevelsens roll i filosofi och humaniora" [On the Role of Imagination in the Human Sciences], in *Att förstå inom humaniora* [Understanding in the Human Sciences]. Helsingfors: Finska Vetenskaps-Societeten.

Hertzberg, L. (2005). "The Limits of Understanding" *Sats-Nordic Journal of Philosophy*, 6, 1: 5-14.

Hertzberg, L. (2005). "Trying to Keep Philosophy Honest", in *Wittgenstein: the Philosopher and his Works*. Edited by A. Pichler and S. Säätelä. Bergen: The Wittgenstein Archives at the University of Bergen.

Hertzberg, L. (2007). "On the Importance of Being Thoughtful", in *Perspicuous Presentations: Essays on*

Wittgenstein's Philosophy of Psychology, edited by D. Moyal-Sharrock. Basingstoke: Palgrave Macmillan.

Hertzberg, L. (2009). "What's in a Smile?", in *Emotion and Understanding: Wittgensteinian Perspectives*, edited by Y. Gustafsson, C. Kronqvist and M. McEachrane. Basingstoke: Palgrave Macmillan.

Hobson, R. P. (1993). *Autism and the Development of Mind*. Hove: Psychology Press.

Hobbes, T. ([1651] 1996). *Leviathan*. Cambridge: Cambridge University Press.

Homans, J. A. (2012). "Tony Judt: A Final Victory" *New York Review of Books*, March 22, 59, 5: 4-7.

Hume, D. ([1739-1740] 1978). *A Treatise of Human Nature*, edited by L.A. Selby-Bigge. Second edition. Oxford: Clarendon Press.

Hurtig, R., Ensrud, S. and Tomblin, J. B. (1982). "The Communicative Function of Question Production in Autistic Children" *Journal of Autism and Developmental Disorders*, 12: 57-69.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C. and Rizzolatti, G. (2005). "Grasping the Intentions of Others with One's own Mirror Neuron System" *PLOS Biology*, March, 3, 3, e79: 0529-0535.

- Judt, T. (2010). "Night" *The New York Review of Books*, January 14, 57, 1: 4.
- Karf, G. (1999). *En junisöndag kvart över tolv* [A Sunday in June a Quarter Past Twelve]. Helsingfors: Söderström & Co.
- Kanner, L. (1943). "Autistic Disturbances of Affective Contact" *Nervous Child*, 2: 217-50.
- Kukla, R. (2005). *Mass Hysteria: Medicine, Culture, and Mothers' Bodies*. Oxford: Rowman & Littlefield.
- Lagerspetz, O. (2008). "Perception and Shared Experiences". Unpublished paper originally presented at the University of Aarhus, 22 May.
- Leder, D. (1990). *The Absent Body*, Chicago: The University of Chicago Press.
- Leudar, I. and Costall, A. (2009). *Against Theory of Mind*, edited by I. Leudar and A. Costall. London: Palgrave Macmillan.
- Levi, D. (2000). *In Defence of Informal Logic*. Dordrecht: Kluwer.
- Levinas, E. ([1961] 1969). *Totality and Infinity*. Pittsburgh: Duquesne University Press.

Lewis, C. S. ([1964] 1991). *Letters to Malcolm Chiefly on Prayer*. London: Harcourt Inc.

Lexhed, J. (2008). *Det räcker inte med kärlek, En mammas kamp för sin son* [Love is not Enough: A Mother's fight for her Son]. Stockholm: Bonnier Pocket.

Lipps, T. (1897). *Raumästhetik und geometrisch-optische Täuschungen*. Leipzig: JA Barth.

Luria, A. R. (1976). *Cognitive Development, its Cultural and Social Foundations*. Cambridge: Harvard University Press.

Løgstrup, K. E. ([1956] 1997). *The Ethical Demand*. Notre Dame: University of Notre Dame Press.

Mahler, M. S., Pine, F. and Bergman, A. (1975). *The Psychological Birth of the Human Infant*. New York: Basic Books.

Malcolm, N. (1972). *Problems of Mind: Descartes to Wittgenstein*. London: George Allen & Unwin LTD.

McHale, S. M., Simeonsson, R.J. , Marcus, L.M. and Olley, J.G. (1980). "The Social and Symbolic Quality of Autistic Children's Communication" *Journal of Autism and Developmental Disorders*, 10: 299-310.

Meltzoff, A. and Moore, M. K. (1977). "Imitation of Facial and Manual Gestures by Human Neonates" *Science, New Series*, 198, 4312, October 7: 75-78.

Meltzoff, A. and Moore, M. K. (1983). "Newborn Infants Imitate Adult Facial Gestures" *Child Development*, 54: 702-709.

Meltzoff, A. and Moore, M. K. (1994). "Imitation, Memory, and the Representation of Persons" *Infant Behavior and Development*, 17: 83-99.

Meltzoff, A. N. and Moore, M. K. (1997). "Explaining Facial Imitation: A Theoretical Model" *Early Development and Parenting*, 6: 179-192.

Meltzoff, A. (2002). "Imitation as a Mechanism of Social Cognition: Origins of Empathy, Theory of Mind, and the Representation of Action" in *Blackwell Handbook of Childhood Cognitive Development*, edited by U. Goswami. Oxford: Blackwell Publishers.

Meltzoff, A. (2010). "Bridging Between Action Representation and Infant Theory of Mind" in *Cognition and Neuropsychology: Vol. 1. International Perspectives on Psychological Science*, edited by P. A. Frensch and R. Schwarzer. New York: Psychology Press.

Meltzoff, A. (2011). "Social Cognition and the origins of Imitation, Empathy and Theory of Mind" in *The Wiley-*

Blackwell Handbook of Childhood Cognitive Development, second edition. Edited by U. Goswami. Malden, MA: Wiley-Blackwell.

Meltzoff, A. (2013). "Origins of Social Cognition: Bidirectional Self-Other Mapping and the Like-Me Hypothesis" in *Navigating the Social World: What Infants, Children and Other Species Can Teach Us*, edited by M. Banaji and S. Gelman. New York: Oxford University Press.

Meltzoff, A., Williamson, R. A. and Marshall, P. J. (2013). "Developmental Perspectives on Action Science: Lessons from Infant Imitation and Cognitive Neuroscience" in *Action Science: Foundations of an Emerging Discipline*. Edited by W. Prinz, M. Beisert and A. Herwig. Cambridge, MA: MIT Press.

Nagel, T. (1974). "What is it Like to be a Bat?" *Philosophical Review*, 83, October: 435-50.

Nussbaum, M. (2001). *Upheavals of Thought*. Cambridge: Cambridge University Press.

O'Neill, D. (1996). "Two-Year-Old Children's Sensitivity to a Parent's Knowledge State When Making Requests" *Child Development*, 67, 2, April: 659-677.

Park, C. C. (1967). *The Siege, a Family's Journey Into the World of an Autistic Child*. Boston: Little, Brown and Company.

Park, C. C. (2001). *Exiting Nirvana: a Daughter's Life with Autism*. Boston: Little, Brown and Company.

Piaget, J. (1952c). *The Origins of Intelligence in Children*. New York: International Universities Press.

Piaget, J. ([1954] 1971). *The Construction of Reality in the Child*. New York: Ballantine books.

Piaget, J. ([1955] 1974). *The Language and Thought of the Child*. New York: The New American Library.

Plügge, H. (1967). *Der Mensch und Sein Leib*. Tübingen: Max Niemeyer Verlag.

Pratt, M. L. (1986). "Fieldwork in Common Places" in *Writing Culture. The Poetics and Politics of Ethnography*, edited by J. Clifford and G. Marcus. Berkeley: University of California Press.

Premack, D. and Woodruff, G. (1978). "Does the Chimpanzee Have a 'Theory of Mind'?" *Behavioral and Brain Sciences*, 4: 515-526.

Prizant, B. and Duchan, J. (1981). "The Functions of Immediate Echolalia in Autistic Children" *Journal of Speech and Hearing Disorders*, 46: 241-249.

Prizant, B. and Rydell, P. (1984). "Analysis of Functions of Delayed Echolalia in Autistic Children" *Journal of Speech and Hearing Research*, 27: 183-192.

Reddy, V. (2008). *How Infants Know Minds*. Cambridge: Harvard University Press.

Reddy, V. and Morris, P. (2009). "Participants Don't Need Theories: Knowing Minds in Engagement" in *Against Theory of Mind*, edited by I. Leudar and A. Costall. London: Palgrave Macmillan.

Ricks, D. M. and Wing, L. (1975). "Language, Communication and the Use of Symbols in Normal and Autistic Children" *Journal of Autism and Childhood Schizophrenia*, 5: 191-122.

Rizzolatti, G. (2005). "The Mirror Neuron System and its Function in Humans" *Anatomy and Embryology*, 210: 419-421.

Rizzolatti, G. and Arbib, M. A. (1998). "Language Within our Grasp" *Trends Neurosci*, 21: 188-194.

Rizzolatti, G. and Craighero, L. (2004). "The Mirror-Neuron System" *The Annual Review of Neuroscience*, 27: 169-192.

Rubin, S., Biklen, D., Kasa-Hendrickson, C., Kluth, P., Cardinal, D. N. and Broderick, A. (2001). "Independence,

Participation, and the Meaning of Intellectual Ability" *Disability and Society*, 16, 3: 421.

Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson's University Library.

Schaffer, R. (1977). *Mothering*. Cambridge: Harvard University Press.

Shanker, S. (2004). "The Roots of Mindblindness" *Theory & Psychology*, 14: 685-703.

Shapiro, T. (1977). "The Quest for a Linguistic Model to Study the Speech of Autistic Children." *Journal of the American Academy of Child Psychiatry*, 16: 608-619.

Shostak, M. (1981). *Nisa: The Life and Words of a !Kung Woman*. Cambridge: Harvard University Press.

Skinner, B. F. (1953). *Science and Human Behavior*. New York: Macmillan.

Smith, J. M. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

Smith, A. ([1759] 2002). *The Theory of Moral Sentiments*. Cambridge: Cambridge University Press.

Sontag, S. (1977). *On Photography*. New York: Farrar, Straus and Giroux.

Sontag, S. (2003). *Regarding the Pain of Others*. New York: Farrar, Straus and Giroux.

Sperber, D. (1993). Paper presented at conference on Darwin and the Human Sciences, London School of economics, June.

Sperber, D. and Wilson, D. (1987). "Précis of Relevance: Communication and Cognition." *Behavioural and Brain Sciences*, 10: 697-710.

Stotland, E. (1969). "Exploratory Investigations of Empathy" in *Advances in Experimental Social Psychology, Vol. 4*, edited by L. Berkowitz. New York: Academic Press.

Stueber, K. R. (2006). *Rediscovering Empathy*. Cambridge: The MIT Press.

Tally, T. (1991). *The Silence of the Lambs*, directed by J. Demme. USA: Orion Pictures.

Tantam, D. J. H. (1991). "Asperger's Syndrome in Adulthood" in *Autism and Asperger Syndrome*, edited by U. Frith. Cambridge: Cambridge University Press.

Thomése, P. F. ([2003] 2005). *Shadow Child*. London: Bloomsbury.

Titchener, E. B. (1909). *Lectures on the Experimental Psychology of Thought Processes*. New York: Macmillan.

Tooby, J. and Cosmides, L. (1997). "Foreword" in S. T. Baron-Cohen *Mindblindness: An Essay on Autism and Theory of Mind*. London: The MIT Press.

Tooby, J, Cosmides, L., Sell, A., Lieberman, D. and Sznycer, D. (2008). "Internal Regulatory Variables and the Design of Human Motivation: a Computational and Evolutionary Approach" in *Handbook of Approach and Avoidance Motivation*, edited by A. Elliot. London: Taylor & Francis Group.

Torrkulla, G. (2009). "Unconcluded Remarks on the Countless Configurations of Smiling" in *Wittgenstein and Philosophical Psychology: Essays in Honour of Lars Hertzberg*, edited by C. Gefwert and O. Lagerspetz. Uppsala: Uppsala Universitet.

Updike, J. (1983). "Pain" *The New Republic*, 26 December: 34.

Wetherby, A. M. (1986). "Ontogeny of Communicative Functions in Autism" *Journal of Autism and Developmental Disorders*, 16: 295-316.

Westermarck, E. (1917). *The Origin and Development of the Moral Ideas, Vol. II*. Second edition. London: Macmillan and Co.

Westermarck, E. (1932). *Ethical Relativity*. London: Kegan Paul, Trench, Trubner & CO. LTD.

Wimmer, H. and Perner, J. (1983). "Beliefs About Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception." *Cognition*, 13: 103–128.

Winch, P. (1997). "Can We Understand Ourselves?" *Philosophical Investigations*, 20, 3: 193–204.

Winch, P. (1958). *The Idea of a Social Science and its Relation to Philosophy*. London: Routledge & Kegan Paul.

Wittgenstein, L. (1980). *Remarks on the Philosophy of Psychology, Vol. II*, edited by G. H. von Wright and H. Nyman. Translated by C. G. Luckhardt and M. A. E. Aue. Oxford: Basil Blackwell.

Wittgenstein, L. ([1953] 2001). *Philosophical Investigations*. Third edition. Translated by G. E. M. Anscombe. Malden: Blackwell.

Wittgenstein, L. ([1967] 1981). *Zettel*. Second edition. Edited by G. E. M. Anscombe and G. H. von Wright. Translated by G. E. M. Anscombe. Oxford: Blackwell Publishers.

Wulff, S. (1985). "The Symbolic and Object Play of Children with Autism: A review", *Journal of Autism and Developmental Disorder*, 15: 139-148.

Zahavi, D. (2005). *Subjectivity and Selfhood*. Cambridge: The MIT Press.

Ylva Gustafsson

Interpersonal Understanding and Theory of Mind

What does it mean to understand someone else? What is altruism? What is empathy? How does a child learn to understand other people? The claim that a “theory of mind” is a fundamental cognitive capacity that grounds human social life is popular within both modern philosophical and psychological theorising on interpersonal understanding. This claim surfaces in evolutionary psychology, in theories of child development, in theories of autism as well as in philosophy on emotions and in moral philosophy. The aim of this work is to scrutinise certain psychological and philosophical theories on interpersonal understanding that are connected with empirical research. The author argues that the theories as well as the empirical research are often based on problematic philosophical assumptions about interpersonal relations.



9 789517 657488 >