

Pessi Lyyra

Higher-Order Theories of
Consciousness
An Appraisal and Application



JYVÄSKYLÄ STUDIES IN EDUCATION, PSYCHOLOGY AND SOCIAL RESEARCH 387

Pessi Lyyra

Higher-Order Theories of Consciousness

An Appraisal and Application

Esitetään Jyväskylän yliopiston yhteiskuntatieteellisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Villa Ranan Blomstedt-salissa
huhtikuun 21. päivänä 2010 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Social Sciences of the University of Jyväskylä,
in the building Villa Rana, Blomstedt hall, on April 21, 2010 at 12 o'clock noon.



UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2010

Higher-Order Theories of Consciousness

An Appraisal and Application

JYVÄSKYLÄ STUDIES IN EDUCATION, PSYCHOLOGY AND SOCIAL RESEARCH 387

Pessi Lyyra

Higher-Order Theories of Consciousness

An Appraisal and Application



UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2010

Editor

Jussi Kotkavirta

Department of Philosophy and Social Sciences, University of Jyväskylä

Pekka Olsbo

Publishing Unit, University Library of Jyväskylä

Cover picture by Pessi Lyyra

URN:ISBN:978-951-39-3919-9
ISBN 978-951-39-3919-9 (PDF)

ISBN 978-951-39-3848-2 (nid.)
ISSN 0075-4625

Copyright © 2010, by University of Jyväskylä

Jyväskylä University Printing House, Jyväskylä 2010

ABSTRACT

Lyyra, Pessi

Higher-order theories of consciousness: An appraisal and application

Jyväskylä: University of Jyväskylä, 2010, 163 pp.

(Jyväskylä Studies in Education, Psychology and Social Research

ISSN 0075-4625; 387)

ISBN 978-951-39-3919-9 (PDF), 978-951-39-3848-2 (nid.)

The subjectively experienced 'phenomenality' of consciousness is often taken to constitute the greatest enigma of the contemporary philosophy of mind, or even science in general: how can phenomenal conscious experience arise out of its physical substrate? The so called higher-order theories of consciousness propose to solve this problem by the cognitive ability of being aware of one's own mental states. The first and foremost justification for these theories is the intuitively appealing idea that only those mental states are conscious, which we are somehow aware of. In this study, these theories and the core intuition they rely on are examined. It is claimed that while this intuition holds for everyday human consciousness, it does not hold equally well for the more primitive kinds of consciousness, such as infant or animal consciousness. The intuition also conflicts with the opposite intuition that self-awareness can be detrimental to the vividness of world-directed consciousness. The intuition itself remains unexplained, and the theories lack relevant empirical support. To solve these problems, a novel conception of higher-order consciousness is suggested, a practical form of consciousness characteristic of human beings called 'mentalized consciousness'. It essentially involves self-awareness and develops in the early interaction of an infant and its primary caregivers, as suggested by recent theories and findings in empirical psychology. It is shown that, while avoiding their problems, the intuitive appeal of the higher-order theories of consciousness can be explained by the concept of mentalized consciousness.

Keywords: Consciousness, Phenomenality, Self-awareness, Higher-order theories of consciousness, Phenomenological theories of consciousness, Mentalization, Mentalized consciousness

Author's address

Pessi Lyyra
Department of Philosophy and Social Sciences
University of Jyväskylä, Finland
P.O.BOX 35, 40014 University of Jyväskylä
pessi.lyyra@jyu.fi

Supervisors

Jussi Kotkavirta
Department of Philosophy and Social Sciences
University of Jyväskylä, Finland

Mikko Yrjönsuuri
Department of Philosophy and Social Sciences
University of Jyväskylä, Finland

Reviewers

Paavo Pylkkänen
Department of Theoretical Philosophy
University of Helsinki, Finland

Panu Raatikainen
Department of Theoretical Philosophy
University of Helsinki, Finland

Opponent

David Rosenthal
Department of Philosophy and Cognitive Science
Graduate Center, City University of New York, USA

ACKNOWLEDGEMENTS

I have received inspiration and support for this project from so many people that only few of them can be duely mentioned. First of all, I would like to thank my supervisors Jussi Kotkavirta and Mikko Yrjönsuuri and other people of the unit of Philosophy in the Department of Philosophy of The University of Jyväskylä for guidance, friendship and making up an exemplary working environment filled with enthusiasm and freedom of thought. It has been a privilege to work in a place with so many people with shared interests. People in the research projects "The Concept of Person", "Limits of Personhood" and "Human and animal minds in European philosophy 1200-1700" have influenced this study the most. In our seminars and reading groups most ideas of this study were conceived, developed and tested during the years. This study has crucially benefitted from discussions on the issues of it with Professor Antti Revonsuo from the University of Turku, Professor Dan Zahavi from the University of Copenhagen, Professor Shaun Gallagher from the University of Orlando, and Dr. Susanna Radovic from the University of Gothenburg. Mr. Michael Freeman gave indispensable advice with the English language.

The research has been supported by The Finnish Cultural Foundation, The Fund of Central Finland of The Finnish Cultural Foundation, respectively, The Academy of Finland, and The Agora Center of The University of Jyväskylä.

The research project has also been a significant occasion of personal growth, of self-understanding and becoming the person I am, which is reflected in the content of the study. I can say that I have lived the best days of my life during this research project. I owe everything to the "strong Lyyra women" and other members of my family, especially to my mother Ulla-Riitta Lyyra. I feel that I have inherited some of her originality and courage to stand behind it. Finally, I wish to thank the most my wife Hannalea Lyyra and our children Nana and Jouka for all the support and love - you have taught me what it is all about.

At Gift Ridge, March 13, 2010
Pessi Lyyra

FIGURES

FIGURE 1 The consciousness tree.....	34
FIGURE 2 The consciousness tree, version 2.....	37
FIGURE 3 The consciousness tree, version 3.....	42
FIGURE 4 The consciousness tree, version 4.....	67
FIGURE 5 The consciousness tree, version 5.....	109

TABLES

TABLE 1 Objections to HO-theories.....	92
--	----

CONTENTS

ABSTRACT

ACKNOWLEDGEMENTS

FIGURES AND TABLES

CONTENTS

1	INTRODUCTION	9
1.1	Consciousness and self-awareness.....	9
1.2	An itinerary.....	13
2	THE PROBLEM OF CONSCIOUSNESS.....	15
2.1	On method: The return of speculative psychology?.....	15
2.2	Consciousness: basic distinctions	22
2.3	The “easy” and “hard” problems of consciousness.....	26
2.4	Empirical approaches to consciousness	34
3	THE REPRESENTATIONAL THEORY OF CONSCIOUSNESS.....	42
3.1	Mind and intentionality	44
3.2	Representational content: narrow or wide.....	50
3.3	Representational content: conceptual or non-conceptual?	57
3.4	Evaluative intentionalism.....	62
4	HIGHER-ORDER THEORIES OF CONSCIOUSNESS.....	67
4.1	First-order and higher-order representationalism: a first approximation.....	67
4.2	Two basic views: HOP and HOT	71
4.3	Intrinsicalism and relationalism	84
4.4	HOR and language	88
4.5	Appraisal: Insufficiency of HOR in explaining phenomenality	92
4.6	HOP and HOT: a false opposition?.....	105
5	FORMS AND PROPERTIES OF HIGHER-ORDER CONSCIOUSNESS .	110
5.1	Objects of higher-order representations.....	110
5.2	Psychological issues in higher-order representations	116
5.3	Mentalized consciousness and other forms of higher-order consciousness	122
5.4	The intuition of “givenness of consciousness”	129
5.5	Contenders and counter-arguments	133
6	CONCLUDING REMARKS	136
	YHTEENVETO.....	142
	REFERENCES.....	147

1 INTRODUCTION

1.1 Consciousness and self-awareness

Two of the most fundamental questions concerning mind are the object of this study. The first is the metaphysical question about phenomenal conscious experience: how can it feel like something rather than nothing to be me? Or, how can phenomenal conscious experience arise out of its physical substrate? This is arguably the most pressing metaphysical question in the modern philosophy of mind. Nonetheless, reflexive epistemological questions have on occasion been considered as even more fundamental than the metaphysical ones. The second fundamental question is the reflexive epistemological question: How do I *know* that it feels like something rather than nothing to be me?

Many have started to wonder whether, in fact, there is any difference between one's phenomenal experience and one's awareness of it. For they believe that there being something it is like to be me is the same as my knowledge of it. This natural way of thinking constitutes one branch of theories of phenomenal conscious experience. David Rosenthal has formulated this general intuition in the most straightforward and explicit way as the "*transitivity principle*" (TP):

(TP): "A mental state is conscious only if one *is in some way aware* of it." (Rosenthal 2005, 4; 2006, 7, emphasis PL; cf. Lycan 2001, 3)

In this study, my task is to judge the validity of this principle and theories based on it. These theories, the so called *higher-order* theories of consciousness are the main topic of this thesis. The term "higher-order" (HO for short) refers, quite simply, to awareness of mental states, whereas "first-order" (FO for short) awareness means here awareness of other things, usually things in the (outside) world. Higher-order theories of consciousness claim that consciousness requires self-awareness in the form of awareness of one's own mental states, while for first-order theories phenomenality is awareness of the world or is based solely on certain kind of neural activation.

My aim is to show that adopting the higher-order view may be a problematic and not the best way to solve the metaphysical question about phenomenal experience. However, I also aim to show that the discussion surrounding the higher-order theories is not futile despite this shortcoming. The intuition supporting the principle is strong and it seems to capture something of what laypersons at times mean by consciousness. Therefore, the potential of these theories deserve a comprehensive evaluation. I will attempt to show that the appeal of the principle is due to the ambiguity inherent in the concept of consciousness, and that the validity of the principle depends on what we mean by consciousness. Different concepts of consciousness should be distinguished, and I will argue on conceptual, phenomenological and developmental grounds that self-awareness is crucial for some but not all meanings of consciousness. I will ultimately argue that the transitivity principle is not valid for the conception of consciousness as phenomenal experience. As a replacement of a kind, I will suggest a novel, mundane sense of object-directed consciousness in which consciousness and self-awareness do indeed converge. The explication of this mundane sense of consciousness in terms of higher-order theories of consciousness is the second, more positive outcome of this thesis. In unraveling this issue, some light will be shed upon the way consciousness and self-awareness are intermingled in general.¹

Philosophers have traditionally considered the ability to be aware of one's mental states as special both epistemologically and psychologically. As an initial approximation, I will explore some of the basic intuitions that philosophers have had about the nature of self-awareness and higher-order representations. These intuitions impose a number of basic criteria on a theory based on higher-order representations. The second aim of the thesis cannot be reached without fulfilling these criteria.

The ability to distinguish between *mind* and *world*, or as developmental psychologists often put it, the ability to distinguish between *appearance* and *reality* seems to be embedded deep in our cognitive capacities². We can consciously attend to the internal mental realm, and observe what is occurring

¹ The second strand concerning the relation between consciousness and self-awareness is not that consciousness is due to awareness of it, but that the very *problem* of consciousness is. Consciousness is hard to explain because our first-hand ways of knowing about phenomenal conscious experience are so detached from scientific concepts. This gives rise to the problem of consciousness. On this basis, some think that this conceptual dualism should extend to metaphysical level. The important question is whether we could rid ourselves of conceptual dualism, if we start to think of consciousness using concepts that can be connected in the relevant way with scientific concepts. Some continue to think that this is impossible, and that the problem lies in the metaphysical nature of phenomenal experience.

² Distinguishing between mind and the world is only one specific version of this distinction. The two most fundamental ways of making this distinction are described by Husserl in the following manner: "We cannot too sharply stress the equivocation which allows us to use the word 'appearance' *both of the experience in which the object's appearing consists [...] and of the object which appears as such.*" (Husserl 1970, 538). There are yet other ways of making this distinction such as when it is applied only to mental states. All these other meanings are irrelevant in this context except the first, the *experience* as appearance of objects distinguished from the objects of the experience.

in it at that time. If we had no knowledge of our own mentality it would be impossible to distinguish between mental states and the world.

Knowledge of mentality is not confined to self-knowledge. In fact, the most common instance of this knowledge is that concerning others rather than oneself. We attribute mental states quite naturally to each other. Our world is at the same time social and psychological as much as it is physical. Much of our every day speech is filled with folk-psychological vocabulary: “I *saw* that he was *hesitating* and I *thought* he’d *be angry* if he wouldn’t have *dared* to do it” or “She *knows* well enough that I don’t *like* garlic and yet *wishes* that I would *learn* to *like* it and is always adding it in my food. She *thinks* she *knows* better than me what I should eat”. Simon Baron-Cohen has observed that folk-psychological concepts are embedded so deep in our understanding of the world that we do not normally consciously realize how abundant and automatic their use actually is (Baron-Cohen 1995). We recognize other persons as mental subjects as much as physical creatures (Strawson 1959). All this constitutes our knowledge of mentality. At the psychological level, the ability to deal with psychological information has been given many names in psychology and philosophy – social cognition, theory of mind, folk psychology, common sense psychology, to name a few. Different accounts exist of it as well as different opinions on how unitary a phenomenon it is.

From these considerations, I wish to draw two criteria that a theory of psychological self-awareness, awareness of one’s own psychological qualities, has to meet. First, it must explain how we can distinguish psychological phenomena from everything else: how do we discriminate and receive psychological information? Secondly, how can we make use of this information and retain it over time? The received view is that this is possible in virtue of a set of psychological concepts, a “theory of mind”. A good theory must include an account of it or of a corresponding capacity.

One’s relation to oneself seems partly mediated by the same psychological concepts as one’s relation to others. In both philosophy and psychology, it has been a popular to view the *access* to oneself as essentially the same as to other persons, since the meaning of psychological concepts is constituted by their use in social circumstances and thus the relation to oneself has to be of the same kind if it is supposed to consist of the same concepts (Sellars 1956; Wittgenstein 1953, Wittgenstein 1974; Gopnik 1993).

However, the more traditional view is that one’s knowledge of oneself fundamentally differs from one’s knowledge of others. There is a difference in knowing that I myself am thinking and knowing that another is thinking, even if the concept of thinking remains the same. This is possible, since one’s mind seems to be readily “given” or present to oneself in a way that allows introspective *access*. Yet, the nature of this givenness is not clear. The givenness itself is not immediately given in introspection, and it is debated whether it can be brought to the level of experiential givenness at all (Zahavi 1998). Put more contrastively, how does access to mind differ in the case of oneself and others? One seems to have a *privileged* or a special access to one’s own mind. Being able to explaining this access forms the third important criterion for any such

theory³. It could even be argued that psychological concepts do not themselves explain one's access to one's mental states, but instead that the access to one's mind enables the formation psychological concepts (Goldman 1993). This issue is not clear-cut, and it may be that access and concepts depend on each other.

In spite of the peculiarity of the access to one's own mind, it has often been put in terms of knowledge of external things. It is generally thought that one can "look" into one's own mind by virtue of an "inner eye" (e.g., Humphrey 1986). This is of course a figure of speech, but the basic intuition still stands that inner perception resembles outer perception in many respects. The elucidation and assessment of the theoretical positions concerning this intuition are major tasks of this thesis. Higher-order theories of consciousness differ with respect to how they define the access of the higher-order state to its target state.

Self-awareness is not an unambiguous term. On the one hand, it can refer to awareness of one's own mental states. On the other hand, it can mean awareness of a *self* (Kriegel 2007). It can be argued that minimal consciousness of our selves is an important part of consciousness (Kriegel 2005). One way to see this is to understand that every mental episode has some kind of *self-aspect* to it. Consciousness is *someone* being conscious. When I reflect my mental states, they appear immediately as mine, not as anyone else's. It has been argued that there must be a self-aspect to mental states for them to constitute *self-awareness* in the first place (Henrich 1970; Zahavi 1998). The third important criterion a good theory of self-awareness needs to satisfy is to spell out the nature of the possibility of self-awareness. Some theories have defined this in terms of awareness of mental states (Zahavi 1998; Frank 1999). I will argue that this is best conceived of in terms of a self-aspect or *indexicality* of mental states.

The fifth and final criterion is accounting for the already mentioned strong appeal of the transitivity principle. Together, these criteria should form the backbone of any theory of metarepresentational consciousness. Although there may be more promising theories that purport to explain phenomenal consciousness, higher-order theories might turn out to be useful in explaining metarepresentational forms of consciousness. The strong intuitive appeal of a higher-order theory despite its failure to solve the problem of phenomenal consciousness is commonly noted. Higher-order theories might be good for explaining some other phenomena related to consciousness, as suggested by both Ned Block (1995) and David Chalmers (1996). This path has not been taken, as far as I am aware, hence this study.

³ Epistemologically, an increased degree of certainty is often taken to follow from the idea of privileged access. I will not consider epistemological questions more than is necessary for present purposes (see Alston [1972] for epistemological considerations). I am more interested in the nature of the psychological capacities behind higher-order consciousness. Epistemological issues may, however, profit from psychological investigation.

1.2 An itinerary

The present study of the higher-order theories of consciousness has a two-fold aim. First, it is a criticism of the higher-order theory of consciousness as a theory of phenomenality of conscious experiences. Attaining the first sets the stage for the second aim, that of applying the higher-order theories in explaining a specific form of consciousness. To this end, three important suggestions are made in this study: a concept of consciousness that the higher-order theories suit to explain, mentalized consciousness, a view of pre-reflective self-awareness in terms of indexicality, and a dispositional view of the givenness of consciousness. I attempt to show how these suggestions enable meeting the criteria introduced in the previous section: 1) the special access to one's own mind 2) possibility to extract, retain and apply psychological information, 3) an account of *whose* access and capacities these are, 4) the possibility of being reflectively aware of one's mental states, and 5) the intuition that one is conscious of one's conscious states.

First, higher-order theories should be presented in their original context, the debate over the problem of consciousness in the modern philosophy of mind and in the cognitive and neural sciences. The problem of consciousness and the status of HO theories cover chapters 2 to 4. My approach to these problems slightly differs in aim and methodology from the mainstream approach. As a preliminary to the problem of consciousness, I first describe the general methodological approach and its basic ontological commitments. Then I review some of the aspects of the problem of consciousness relevant for the assessment of HO theories as a solution to the problems surrounding consciousness. In addition, chapter 2 introduces some basic terminology related to consciousness. Chapter 3 presents the wider context to which the HO theories belong, that of *representational* theories of consciousness. Chapter 4 is devoted specifically to the various forms and aspects of the higher-order theories, especially as theories of consciousness. At the end of chapter 4 a number of specific criticisms of the higher-order theory are presented. These pave the way for the application of the theory, which is the aim of chapter 5. In this chapter, I apply the higher-order theories in distinguishing different forms of psychological self-awareness and the psychological capacities required by them. The most important outcome is that higher-order theories can clarify an important aspect of consciousness that has not been the usual target in consciousness studies. This, what I like to call 'mentalized consciousness', is a sort of world-directed consciousness that is guided by psychological information and has a personal history in one's psychological development in interaction with others that have the same abilities. One has to be brought up to be conscious of and gain control over the events of one's actions and inner mental life.

I close by drawing some preliminary conclusions about the results of this study and their future applications.

Themes of this study have previously been dealt with in Lyyra (2004; 2005a; 2005b & 2009), and some conference presentations and Finnish papers. The author wishes to thank the publishers for the permission to reproduce some of this previously published material.

2 THE PROBLEM OF CONSCIOUSNESS

Although consciousness has not been investigated systematically until since 1970's⁴, the amount of the literature on the problem of consciousness has rapidly accumulated to the extent that it is virtually impossible to take into account every strand of the recent research done to solve the problem in different scientific disciplines. My aim is neither to solve the problem, nor even to attempt to follow comprehensively every ramification of the discussion. For the purposes of this study, it suffices to present some of the issues that are relevant for the assessment of the higher-order theories as theories of consciousness. It will be difficult to remain neutral to the problem, and I will refer to some of the ideas that I see as promising for a possible solution to the problem of phenomenal consciousness.

In this chapter, I first describe my general strategy, which utilizes the results of empirical sciences in guiding and constraining conceptual considerations (cf. Bermúdez 1998; Papineau 1993). I then present the most important uses of the term consciousness in the context of this study. Third, I briefly outline the forms of the problems associated with them. Fourth, I review some of the proposed solutions to these problems.

2.1 On method: The return of speculative psychology?

Before tackling the problem of consciousness, I will briefly tutor the reader in the general strategy, the methodological guidelines and the metaphysical commitments of my approach. This is because some of the ideas guiding this work have been presented relatively recently, and thus deviate from the mainstream approaches in the philosophy of mind. These deviations make up

⁴ However, during the period of introspectionism around the turn of 19th and 20th centuries consciousness was defined as the object of psychological research (see especially James 1890). Also phenomenology can also be seen as the systematic description of consciousness, although it is more epistemologically concerned than the more psychological and metaphysical approaches to consciousness.

the source of most of the criticisms of HO theories and suggestions for solutions for different problems of consciousness.

The study of consciousness and self-awareness is necessarily multi-disciplinary. It has long been a commonplace to note this, but the reasons for so doing are less commonly explored. One reason for taking into account the points of view of different disciplines is no doubt practical: the scientific practice of study of consciousness has shown that studying only one branch of science is simply not sufficient. It is useful to consult philosophy, the cognitive sciences, the developmental sciences and the neurosciences for guidance and extra resources. It may be possible to restrict oneself to conducting philosophical analyses of consciousness in one's armchair; and some still think that this is the way philosophers should work. Many, however, would not agree with this, and it can be questioned whether it is possible to neglect what is meanwhile being found in laboratories. It is difficult to see how in his analyses a philosopher could avoid appealing in any way to the findings of other disciplines (Papineau 1993, ch. 1). The concepts of natural language are not pre-given or immune to new discoveries or conceptual change. Conceptual change is often triggered by findings in the natural sciences.

In addition to these practical considerations, there are other reasons for adopting a multidisciplinary approach. Multidisciplinary approaches to consciousness are not historically novel. One historical approach was *speculative psychology*, which prevailed during the era of introspectionism, and was advocated by such figures as William James (1891). Jerry Fodor (1975), pleading for allegiance to the strategic tenets of speculative psychology, describes it as follows:

"There used to be a discipline called speculative psychology. It wasn't quite philosophy because it was concerned with empirical theory construction. It wasn't quite psychology because it wasn't an experimental science. But it was dedicated to the notion that scientific theories should be both conceptually disciplined and empirically constrained. What speculative psychologists did is this: They thought about such data as were available about mental processes, and they thought about first-order psychological theories as had been proposed to account for the data. They then tried to elucidate the general conception of the mind that was implicit in the data and the theories." (Fodor 1975, x)

The epistemological climate of the time, however, brought the entire discipline to a halt. At the turn of the 19th and 20th centuries, philosophers saw themselves threatened by "psychologism". Conceptual questions were seen to be solvable by psychological rather than philosophical analysis. This tendency seemed threatening to some celebrated philosophers, such as Gottlob Frege and Edmund Husserl. In consequence, they insisted on strict separation of conceptual analysis and phenomenological research from psychology or any other empirical science. Ever since, analytic philosophers have felt that their purpose is to analyze and describe concepts and, at most, to point to inconsistencies in their meaning or use. This general spirit was attenuated when Quine (1952) famously called into question the distinction between purely

conceptual and purely empirical knowledge. Philosophers inspired by Wittgenstein have nevertheless kept this spirit alive by insisting that the job of philosophy is to evaluate the conceptual clarity of words by examining how words are used, and this should precede or be separated from empirical studies (see, e.g., Bennett & Hacker 2003).

Put in terms more relevant to the present object of study, separating philosophy as analysis of natural language and science means also embracing a *semantic dualism* concerning the personal–subpersonal levels: the discourse describing the personal level is incommensurable with the scientific discourses describing phenomena at the subpersonal levels⁵. Except in the important task that philosophy has in achieving conceptual clarity, I believe that this line of thought is fundamentally mistaken and leads to conceptual stagnation and prevents progress in the sciences of the mind. The return of the spirit of speculative psychology, collaboration of empirical and conceptual investigations in consciousness studies, reflects the fact that this form of philosophy is insufficient in practice.

The view that sees personal and sub-personal levels as incommensurable, like that of Wittgensteinians, is often called an *autonomist* view (Bermúdez 2005): it holds that regularities holding at the personal level cannot possibly be mapped onto those holding at the subpersonal levels. Entities and phenomena at the personal level are individuated in terms that have no application at more basic levels. Good examples of this are normativity and phenomenal experience. In order to understand mind at the personal level, it is futile to turn to the sciences operating at the subpersonal levels such as neuroscience. The neural level is merely a condition of possibility for the personal level. The autonomist view admits of no systematicity across levels.

There are three other ways to react to the problem of the relation between the personal and subpersonal levels as described by Bermúdez (*ibid.*). These other approaches do admit of systematicity across levels. Two of these are *functionalist* positions that operate top-down, starting from a functionalist analysis, a description of a whole system, and a subsequent *decomposition* of it into multiple components that are determined by their functional role in the overall system. Bermúdez distinguishes between philosophical and psychological functionalism: philosophical functionalism starts with a conceptual analysis of pre-given folk concepts. This yields an initial functional description that allows subsequent exploration of the realizations of functions at subpersonal levels (Chalmers & Jackson 2001). This account reserves an independent place for philosophy and conceptual analysis while seeing conceptual analysis as an avenue to the subpersonal levels.

For psychological functionalists, in contrast, how the functions in the system should be characterized is an empirical question, a task to be accomplished by the cognitive sciences. Moreover, the function/realizer relation can also be applied to realizers at levels immediately beneath the ones at which the functions were originally described; realizers of these functions are

⁵ By personal–subpersonal distinction I understand a whole–part relation.

characterized as having their own cognitive function that has realizers at the next lowest level. Description starts from more intelligent systems and decomposes them to ever less intelligent subsystems. Both functionalist positions strongly commit themselves to across-level systematicity that can be discovered by a top-down decompositional analysis.

The third alternative approach described by Bermúdez to the autonomist position is a *co-evolutionarist* position that, in addition to the top-down strategy, believes in bottom-up influences: advances made at the lower levels can bear on descriptions at higher-levels. Advances in neurosciences may affect our understanding of the mind. This approach has been famously carried to the extreme by the Churchlands (Churchland 1986; Churchland 1995) who hold a reductionist eliminativist position according to which in the course of the progress of the neurosciences our folk-psychological vocabulary will ultimately be replaced by a neuroscientific one. This is, fortunately, only one version of the co-evolutionary strategy. More modest and more fruitful ones would only state that advances in neurosciences affect our present psychological concepts and lead them to develop into better ones. Sciences at different levels advance by imposing constraints on each other, in both directions (Craver 2005; see also Varela 1996; Lutz et al. 2002). Versions of this approach go beyond pure functionalism in that they may allow realizers to affect the functional description, while retaining the idea of functionalism in a broad sense. Mere functional description is insufficient; it is essential to explore the realizing *mechanisms*, also to understand the functional level (Craver 2005 & 2007).

The non-autonomist positions that believe in systematicity across levels can also be mapped in slightly different ways, e.g. into systematicist and fundamentalist (reductionist) approaches. They differ from each other in their ontological and commitments and explanatory strategies. For a fundamentalist there is a basic level to which the other levels can be eventually reduced. Systematicists are non-reductionists who believe in the explanatory relevance of multiple levels. Functionalism and modest co-evolutionarism belong to the systematicist tradition while eliminativism and reductionism are forms of fundamentalism. The position that I will defend and represent here is a systematicist co-evolutionarist position that believes in multiple levels, both in terms of ontology and explanatory relevance. A proper understanding of phenomena at the personal level, like consciousness and self-awareness, is constituted by scientific knowledge from different levels, each pertaining to the other. Advances made at different levels bring about progress at other levels as well as contributing to understanding of the whole. This merits the status of naturalism in that it conforms to the way scientific explanation is taken to proceed in the natural sciences, in biology and in the cognitive sciences, and possibly all the way down to physics (Craver 2007).

This position is close to the one Bermúdez himself has defended in various connections on the basis of his views on *inseparability of syntax and semantics* (Bermúdez 1995a; Bermúdez 1995b; Bermúdez 2000; Bermúdez 2003). In the article "Syntax, Semantics and the Levels of Explanation" (1995a) Bermúdez criticizes the autonomist position advocated by John McDowell in his paper

“The Content of Perceptual Experience” (1994a). In that paper McDowell embraces semantic dualism, a strict separation of the personal and sub-personal levels of explanation. He confines syntactic aspects to the subpersonal and semantic aspects to the personal level and insists that only confusion will result if folk psychological concepts are related to those of the cognitive sciences⁶.

Bermúdez shows that we cannot hope for such a strict separation: syntax is inherently dependent on semantics. Put in terms of psychology, the workings of the brain depend on happenings at the psychological level. Bermúdez invites the reader to consider various neuropsychological conditions: “If we hold to the explanatory autonomy of the personal level in neuropsychological cases, what goes on in such cases will be completely incomprehensible” (Bermúdez 1995a, 366). There must be some systematicity across the levels.

As a more general lesson, analysis conducted at the level of syntax (subpersonal level) cannot be separated from analysis conducted at the level of semantics (personal level), or in yet other words, the separation of (logical or conceptual) structure from content is not warranted. Let us suppose that traditional cognitive science is right in positing that thinking is a causal phenomenon in which the syntactic roles of representations in the causal chain are determined by their physical forms. However, according to Saariluoma (1997), the chain will eventually reach the point at which several possible representations exist that could become the next part of the causal chain. In these cases, Saariluoma argues, mere syntactic forms of representation cannot suffice for determining which representation will be selected. In this case, it is necessary to consult the *content* of the representation. And individuating the content brings in the issue of the relation between the representation and the object of the representation. A purely causal, non-semantic or non-intentional account of thinking cannot be satisfactory. Bermúdez takes the same idea to hold for the personal–subpersonal distinction (2000). Events at the personal level constrain which of possible events occur at the subpersonal levels. At the level of science, the personal level constrains what is taken as object of study at the subpersonal level.

These considerations form the basis of the philosophical strategy that deviates from traditional conceptual analysis, a strategy Bermúdez calls “philosophical naturalism”. The strategy is not naturalism in the traditional sense aiming to a reductive explanation (*pace* original coining of the term by [Papineau 1993]). “Naturalism” means here that theory construction should operate equally with the theories and empirical results of the natural sciences as with the analysis of common sense concepts. By the attribute “philosophical” in philosophical naturalism Bermúdez means that the nature of the investigation is primarily conceptual, although not in the traditional sense of seeking

⁶ McDowell’s own target is Daniel Dennett’s theory of consciousness as a subset of subpersonal representational states, linguistic ones that are selected to a sort of “public relations department” whose contents are in the use of other “departments” or specialized processors (Dennett, 1978). Dennett’s theory is a version of a higher-order theory and will be dealt with in section 4.4.

necessary and sufficient conditions or transcendental arguments. Transcendental arguments should be put to the test of reality:

“Transcendental arguments, it has often been objected, cannot tell us anything about the relevant conceptual abilities themselves. All that they can tell us are the implications of our beliefs and intuitions about those conceptual abilities – this, it is claimed, is all that conceptual analysis can be expected to do. Against this, however, I would suggest that thinking through the implications of the neuropsychological cases enables us to put pressure on and modify transcendental arguments so that the arguments we eventually come up with will hopefully be more than just reflections of our beliefs.” (Bermúdez 1995b, 379)

In my view, this is important, insofar as empirical science often merely operates in and results in a conceptual mess. The positive role of philosophy is to enable *conceptual clarity* in theory formation. Conceptual clarity means not only logical coherence but also something else, such as taking into account a number of conceptual difficulties familiar from the history of philosophy, and forming a coherent whole on this basis from different experimental findings.

Adherence to *semantic monism*, the conceptual relatedness of mind and body, as a consequence of the co-evolutionary strategy raises a number of questions. Analytic philosophers have argued that people are “intuitive dualists” about consciousness: concepts concerning phenomenal consciousness are disconnected from concepts concerning material world (Papineau 2002). If semantic monism is true, how is it possible to overcome initial intuitive dualism? There are broadly three ways to conceive of semantic monism. The first is fundamentalism in the form of reductionism or eliminativism, where pre-scientific psychological terms are translated into or replaced by a scientific vocabulary. The second, less dramatic, way is to think of vocabularies as relatively autonomous but as affecting or constraining each other; they are integrated in the same theory and are thus indispensable parts of it. This is the modest co-evolutionarist strategy that I adhere to in this study. The last alternative is that some intermediary discourse is created to mediate originally distinct levels. An example of this strategy is given in (Gallagher 2005) who seeks to formulate body-related terms that are situated between mind and body. Such concepts figure at both original levels, thus establishing a link between levels. I do not deny that these kinds of terms may exist, and the last two versions of semantic monism may not rule each other out.

Semantic monism and philosophical naturalism are not metaphysical views in the full sense, and metaphysical issues and commitments are dealt with in section 2.3. These commitments correspond well to an explanatory strategy that has gained ground in the philosophy of neuroscience, called the mechanistic tradition (Craver 2005 & 2007; Bechtel 2008). Mechanistic tradition is inherently multilevel, in contrast to traditional fundamentalist reductionism. However, it shares with reductionism the idea that things can be understood and explained in terms of their parts. In contrast, there is no one fundamental level as in reductionism, instead all levels are indispensable parts of the multilevel picture. Moreover, parts of the whole are operational parts and

organized in ways that make up the whole and its behavior. In this way the mechanistic tradition incorporates both the causal and constitutive aspects of explanation and is compatible with the modest co-evolutionary strategy.

Another important aspect of mechanistic explanations is that mechanisms need not be described in terms of a set of sentences. Instead, *visualizations* such as images and diagrams can substitute for linguistic descriptions. Diagrams of mechanisms make it possible to *see* how something works, e.g. in the brain. This carries significant psychological plausibility and describes better how explanation works in the practice of neuroscience. I will try to show how adopting this explanatory strategy opens new perspectives on problems in the philosophy of mind which has thus far mostly been occupied with conceptual analysis (see Revonsuo 2006 for a more thoroughgoing proposal). It seems to me that the mechanistic tradition describes well the key aspects of the empirical branches of contemporary consciousness studies, or the “Consciousness Industry” to borrow the term used by Lagerspetz (2002), that have not been acknowledged or explicitly articulated.

As an extra component I wish to emphasize the importance of testimony of introspection and phenomenology⁷. This has become the received view in spite of the fact that we do not have proper theories of introspection (Vermesch 1999). Ignoring the testimony of introspection on that ground would be the same as ignoring the testimony of sight because we do not as yet have sufficient knowledge of the mechanisms of sight. There are also epistemological worries, which stem from the old dogmas of empiricism, that empirical knowledge should be accessible by anyone, and more recent attacks based on experimental data (Nisbett & Wilson 1977). These criticisms are based on dubious epistemological and metaphysical commitments and have lost ground during the last decades. Despite such epistemological worries, introspection is taken to be indispensable, and it has even been suggested that our psychological concepts derive from introspective access to our own minds (Goldman 1993). The advantage of the present study is that it addresses the question of what cognitive and neural mechanisms underlie introspection. Thus, it at least articulates what introspection means in this context. It would indeed be a serious drawback for a theory of mental self-knowledge if it was unable to use the direct testimony of introspection or elucidate what introspection consists of. Together with the use of introspection as a methodological tool, my approach resembles that of the “neurophenomenology” advocated by Varela (e.g. Varela 1996) and his followers.

⁷ Whether by phenomenology it is meant folk psychological phenomenology, as in analytical philosophy, or the continental, more methodologically restricted phenomenology (Zahavi, 1998) is not always clear. Yet other, more empirically inclined versions of phenomenological methodologies have been suggested (Revonsuo 2006). The *status quo* of the first-person phenomenological methodologies is confused, though they are in general accepted as an elementary part of experimental psychology (see especially the articles compiled in Jack & Roepstorff [2003] and [2004]). I cannot settle these disputes here, and instead I simply apply elements of folk psychology, transcendental and empirical phenomenology whenever any of these seem fit.

In sum, I approach the problem of consciousness and the higher-order theories not solely from a conceptual point of view but draw on multiple sources at multiple levels. I lean on empirical psychology in theory formation, and my suggestion for a novel form of consciousness is largely informed and inspired by developmental psychology. The role of philosophy lies in the attempt to maintain conceptual clarity in theory formation. Theory formation usually takes the form of looking for the best explanation. Abductive inferences like this are sensitive to the restrictions imposed by experimental data on which one starts theorizing, and they are attempts to maximize conceptual clarity. They also carry an advantage over the mere correlative strategies of empirical sciences. Model building and theorizing are important and irreplaceable counter-parts to empirical research. This transcends the strict empiricist standpoint held by the natural sciences, and accords theory formation its due position, as in traditional speculative psychology. This is the way consciousness is studied in contemporary consciousness studies, and so it could be maintained that the multi-disciplinary spirit of Jamesian speculative psychology has truly returned in the modern study of consciousness.

2.2 Consciousness: basic distinctions

“Consciousness” is a notoriously ambiguous term. It has been pejoratively called a “mongrel” of different meanings, and confusions about these different meanings have been claimed to be a source of many systematic misunderstandings (Block 1995). Generally, this has become the received view; it no longer suffices to say that one is dealing with the question of consciousness but one has to make clear which of the meanings of consciousness one is dealing with. The following distinctions have gained relative popularity, and in large part constitute the conceptual framework within which the present study approaches the questions of consciousness and self-awareness.

David Rosenthal has distinguished two basic uses of the term consciousness (e.g., Rosenthal 1993b & 1997). The first is “creature consciousness”, consciousness as a property of conscious beings, as in the case of conscious as opposed to non-conscious human beings such as coma-patients. Creature consciousness is a less controversial meaning of consciousness and corresponds pretty much to a creature being awake. Being awake may involve non-conscious mentality and some creatures, which we might hesitate to ascribe consciousness to, like fish, do exhibit wake-sleep cycles. The opposite of creature consciousness should not be sleep, since at least human beings are clearly conscious in some phases of REM or even non-REM sleep: EEGs show the same γ -frequency activation that is the most compelling index of focused consciousness in sleep as in wakefulness. It has been proposed that dreaming opens a fruitful avenue to the study of consciousness (Revonsuo 1995 & 2006).

The second important meaning of consciousness is “state consciousness”, consciousness as the property of *mental states*. A mental state is either conscious or non-conscious – or perhaps unconscious in the Freudian sense. The higher-order theories attempt to explain state consciousness by consciousness of the mental state.

Consciousness can also be conceived of as “transitive” consciousness as opposed to “intransitive” consciousness, according to a further distinction made by Rosenthal (e.g., 1993b & 1997). Consciousness is transitive when it has an object or it is *intentional*. Consciousness is consciousness of something. Correspondingly, consciousness is intransitive when it has no object or its object is not relevant. The difference between the terms is largely one of aspect. The intransitive use corresponds largely to the concept of creature consciousness that also often ignores the relation of consciousness to an object.

The most important distinction concerning the term consciousness pertains to that between conscious states. The distinction is between “access consciousness” and “phenomenal consciousness”, terms coined by Ned Block (1995). Block introduced this originally as a conceptual distinction but it has tended to live a life of its own ever since, especially in the hands of empirical scientist. It could thus be argued that some co-evolution has truly occurred with respect to these concepts of consciousness (see, e.g., Lamme 2003; Koivisto & Revonsuo 2003). I will start with Block’s initial treatment (1995).

By phenomenal consciousness Block means the *experiential* aspect of consciousness, what it is *like* to live through certain experiences (cf., Nagel 1974). Phenomenal consciousness is often called “subjective consciousness” or “qualitative consciousness”, or simply *qualia*⁸. Commonly cited examples of qualia include such things as a stinging toothache, the redness of a tomato, the bitterness of a freshly bitten lemon and the like. Most often referred to by qualia are experiential aspects that are intrinsic to the experience, non-representational and ineffable, not possible to bring to language use as such (Lewis 1929). This radical definition of qualia is best kept apart from that of phenomenal consciousness which is often taken to involve other aspects, such as those of the overall structure of experiences (Van Gulick 2004a) and even the representational properties of external objects. Many problems concerning consciousness involve this radical definition of qualia. However, not all see qualia in this way, most notably *representationalists*. Chapter 3 is largely dedicated to this issue.

Mere subjective experientiality is not quite faithful to the etymology of the term consciousness. Block has conveniently changed his vocabulary from

⁸ *Qualia* is a plural term for the relative and interrogative pronoun *quale*, meaning “what...like?”. The classical description of C. I. Lewis goes like this: “There are recognizable qualitative characters of the given, which may be repeated in different experiences, and are thus a sort of universals; I call these “qualia.” But although such qualia are universals, in the sense of being recognized from one to another experience, they must be distinguished from the properties of objects. Confusion of these two is characteristic of many historical conceptions, as well as of current essence-theories. The quale is directly intuited, given, and is not the subject of any possible error because it is purely subjective” (Lewis 1929, 121).

phenomenal consciousness to “phenomenality” (Block 2001). I will largely follow Block in this and use mainly his later terminology, since I think there are forms of consciousness that better deserve the label of “consciousness”. “Access” consciousness is perhaps one example of it.

By access consciousness Block means mentality characterized solely by its *functional* properties. The definitions of the functional properties of mental states depend on which roles they occupy in the explanation of action. Block defines access consciousness as consciousness of information at the level that is available for theoretical and practical reasoning and the planning of action or production of speech⁹. A *belief* is often held to be a paradigm example of access consciousness. It is debatable if a belief does or does not have any phenomenology specific to it¹⁰. Nevertheless, it is maintained, access consciousness does not, at least conceptually, have to have any distinctive *feel* to it. The determining feature of a belief is not the phenomenology associated with it. Insofar as these two kinds of consciousness are distinct on a conceptual level, it is conceivable that zombies, without phenomenal consciousness, can nevertheless enjoy a full range of access consciousness. It is logically possible, it is often maintained, that it is not *like* anything to be in an access conscious state.

This, rather common view is not, however, shared by everyone in the field. Others have insisted that phenomenal consciousness is more basic and perhaps access is based on there being phenomenal consciousness (e.g., Revonsuo 2006; cf. Baars 1988). It is a well known fact that there is more to phenomenality than we can access (e.g. Block 2002). As an addendum, it is maintained that there is something it is like to have abstract beliefs (Pitt 2004), and that phenomenal consciousness is a prerequisite of other forms of consciousness. Therefore, it is perhaps best to follow Revonsuo (2006) in leaving access consciousness as something that refers to a conceptual construction associated with functionalism. Nevertheless, he calls “real” access-related aspects of consciousness “reflective consciousness”. Reflective consciousness is a form of consciousness requiring focused attention and conceptual capacities, used in theoretical and practical reasoning and the planning of action or production of speech etc., but which is subordinate to or subspecies of phenomenal consciousness. Block himself has ever since taken a more empirical approach to the distinction and maintained that the two kinds of consciousness have distinct neural mechanisms (Block 2005).

It was stressed in the introduction to the present study that an intimate relation exists between consciousness and self-awareness. Self-consciousness is sometimes seen as a special and controversial form of consciousness which, among others, William Lycan “would not touch [...] for a free week on Maui with champagne thrown on it” (Lycan 1996, 5). Like consciousness, self-consciousness is an ambiguous term. It could mean consciousness of a self, a self being conscious of some of its parts, consciousness of a self-concept,

⁹ If one wishes to differentiate between speech and action (see, Austin 1962).

¹⁰ This is the topic of the famous controversy among the introspectionists Titchener and Külpe.

consciousness involving a self-aspect or a combination of these. The first, consciousness of a self, is the most controversial; according to the Humean “elusiveness thesis” states, there should be no consciousness of a self as such, at least not in a quasi-perceptive mode of object-perception (Bermúdez 2003; Metzinger 2004). Others are less controversial. Ultimately, the nature of self-awareness and its relation to a number of different forms of consciousness cannot be duly accounted for until the end of the present study.

The kinds of self-awareness relevant for this study are not those dealing with self as an object. Instead, the relevant ones are those pertaining to our knowledge of our mental states, not knowledge of a self (even in the sense of a self having mental states). An often-mentioned concept of consciousness with this content is that of “monitoring” or “*reflexive*” consciousness, as further distinguished by Block (1995). This could be taken to include all sorts of consciousness that one can have of one’s consciousness, be it direct or indirect, occurrent or standing. I will refer to reflexive consciousness as an occurrent state. This is the form of consciousness that involves self-awareness that has been a topic in the recent consciousness studies. However, I believe that consciousness as something involving self-awareness and especially self-control, which is clearly the meaning of consciousness in the natural use of language, is not properly characterized by occurrent reflexive consciousness. I will claim on the way toward the end of the thesis that a standing, dispositional form exists that constitutes an important form of consciousness in its own right, that of *mentalized* consciousness.

A good example of dormant and occurrent mental states can be given using belief. On one hand, a belief can be dormant, a standing state. Then it merely states what the subject is *disposed* to think, e.g., what the content of his belief would be if he were to consider the existence of qualia¹¹. These kinds of dormant states affect one’s actions and the contents of one’s occurrent mental states. On the other hand, believing could be an *occurrent* or an active state of consciousness that the subject is having at the moment.

A concept related to consciousness and frequently seen as essential for consciousness is that of *attention*. As latent in the previous discussions, phenomenal consciousness is something that does not necessarily require (focused) attention. For the sake of clarity and later purposes, it should be mentioned that attention itself is not a unitary phenomenon. An important distinction exists between voluntary and involuntary attention. One can direct one’s attention toward things but things can also summon one’s attention. Involuntary attention is often seen as implicit and automatic and something that requires no attention. Thus, implicit attention can be seen something prior to or even constitutive of phenomenal consciousness. Implicit attention itself admits of a further distinction between *spatial-selective* attention and *object-selective* attention (Herrmann & Knight 2001). These seem to be forms of attention related to the “two streams” of visual processing (Milner & Goodale 1995), one related to action and spatial aspects of perception and the other to

¹¹ See next section.

recognition and non-spatial aspects of perception. It is not quite clear whether these are genuine forms of attention or whether there is a single attentional system that operates with these two closely linked systems of perception. Be that as it may, it is often useful to keep these forms of attention, at least conceptually, apart. Often lumped together, the concepts of implicit and explicit or involuntary and voluntary attention do not much figure in the debate over the nature of consciousness. However, they will play a part in when it comes to studying the relation of consciousness and self-awareness and my suggestions of mentalized consciousness as a genuine form of consciousness.

Recently, it has been argued that phenomenal consciousness is something that requires no focal attention, unlike attention-dependent access or reflective consciousness (Lamme 2003; Koivisto et al. 2007). Our most distinctive ability to access the contents of mind is arguably by paying attention to them¹².

The list of the meanings of consciousness is not exhaustive, and the meanings of these concepts will hopefully become clearer, and even evolve, in the course of the thesis. Some other concepts may be introduced but I will attempt to interpret them in terms of those described here.

2.3 The “easy” and “hard” problems of consciousness

The problem of consciousness itself has different dimensions, depending on which meaning of consciousness is at stake. The most famous distinction made concerning the problem of consciousness is that made by David Chalmers (1996). His book *The Conscious Mind: In Search of a Fundamental Theory* has already become the *locus classicus* of the problem of consciousness as it is nowadays taken to be. Chalmers distinguishes between two problems of consciousness: an “easy” one and a “hard” one. The ultimate question for both is whether we can explain consciousness in naturalistic terms. For Chalmers, naturalism means the possibility of giving a *reductive explanation* of a phenomenon in terms of lower-level physical phenomena. It should be shown that physical properties determine or entail phenomenal properties. By reductive explanation Chalmers does not exactly mean classical reduction, in which the phenomenon at the upper-level is identified by appropriate bridge principles to the phenomena at the lower-level (or derived from them). Instead of identity, he describes the notion of *supervenience* to be sufficient for reductive explanation. Supervenience refers to a relation of co-variance and dependence between phenomena at two distinct levels. Chalmers explicates this in terms of two sets of properties at distinct levels, A-properties and B-properties (Chalmers 1996, 33). A-properties are lower-level properties on which upper-level B-properties are supervenient. In other words, there could be no change in

¹² There may be ways of accessing that require no consciousness. But then the question is no longer one about access consciousness. Hence, this particular objection is not a good one for doubting the existence of access consciousness.

the B-properties without a change in the A-properties. The concept of supervenience has the advantage over the concept of identity that it permits *multiple realizability*¹³. This means that several kinds of realization could exist at the lower level without any change in the realized properties. It would make no difference to the painting if some bits of its materials are replaced with some other but qualitatively identical ones. The notion of type identity would not permit this.

Chalmers differentiates the concept of supervenience into a few subspecies largely following the distinctions made famous by Kim (1993). If supervenience holds only for a given individual or a particular situation or location in a world, supervenience is local. If it holds across all individuals in a given world, then we have *global* supervenience. These kinds of supervenience that hold in a given actual world are termed weak supervenience. Local supervenience is clearly not of use for solving the hard problem. Global supervenience seems more promising, but it allows for the situation of two individuals in a given world who are physically identical but one is phenomenally conscious while the other is not (Kim 1987). Global supervenience needs to be complemented to serve explanatory purposes.

Supervenience is “strong” when it holds across all “possible worlds”. Strong supervenience is related to the concept of logical supervenience, that holds of conceptual necessity. In terms of possible worlds semantics, there is no possible world where B-properties could change or cease to exist without a change in A-properties. This also means that B-properties are entailed by A-properties – because that is the only possible way. But there is also a type of supervenience that is not logical supervenience. This is metaphysical supervenience as mere correlation or co-variance; two properties just co-vary perfectly across all worlds. Chalmers holds that this would allow us to give a reductive explanation where physical properties necessarily determine phenomenal properties at the conceptual level. The way Chalmers presents the concept of logical supervenience closely knits together conceivability and metaphysical possibility. Logical conceivability entails metaphysical possibility, a possible world where such circumstances could actually obtain.

In the case of a reductive explanation, supervenience is something that has to hold of global logical necessity. Otherwise, according to Chalmers, it is conceptually possible that the lower-level physical basis exists without the upper-level phenomenon. In this case, there would be no genuine gain to our

¹³ For token identity this is not so much a problem since each token table would be identical to its material substrate. Nevertheless, it is possible to render token identity as problematic if multiple realizability is extended to pertain to each temporal moment or to modal properties. Token identity thesis states that for each mental token event there is one token physical event that it is identical to. It allows of no counterfactual situation where a single atom has been changed resulting in a different physical event, a notion which runs counter to many intuitions (Baker 1995). Neither is the idea of token identity able to contribute to the systematicity which often obtains between the multiply realizable tokens. It is notable that the concept of supervenience is more flexible with respect to temporal multiple realizability, but it is equally unable to account for the systematicities between the properties across levels.

understanding and there would be no genuine explanation. The arguments of Chalmers for the irreducibility of consciousness are based on this view of reduction that there is a necessary logical link between the concepts pertaining to the upper and lower levels that can be explicated by supervenience. A necessary conceptual link is the requirement for reductive explanation. Without conceptual entailment, explanation fails.

Functionalism is something that Chalmers approves of when it comes to the idea of reductive explanation. Although functionalism itself is a very flexible means for theorizing – it works equally well for idealism as for physicalism, for example in the case of mentality, the most important thing is that functionalism lends itself easily to the purposes of the reductivist physicalism. According to Chalmers, a functional explanation is a how-possible explanation in the form of causation. At first, the causal chain consists of abstract mental concepts. The mental concepts are characterized by their causal roles that can be further elucidated by elucidating which properties of the organism underlie the mental capacities referred to by the mental concepts. The elucidation itself is in practice seldom realizable due to our ignorance of the physiology of the organism in question. Fortunately, according to Chalmers, it suffices that there is the in-principle possibility of a physiological explanation that follows straightforwardly from the nature of the mental concepts¹⁴. Functionalism is thus appropriate for reductive explanations.

If we could show that consciousness could take place without its physical or functional basis, or that the physical or functional basis of consciousness could take place without there being consciousness, no explanation of consciousness would be possible. And this is exactly what Chalmers maintains in the case of the hard problem of consciousness but not in the case of the easy problem of consciousness.

The easy and hard problems of consciousness correspond to the different meanings of the term consciousness introduced above. The easy problem corresponds to functional or access consciousness, and the hard problem corresponds to phenomenal consciousness¹⁵. Access or functional consciousness is *per definitio* easily explained functionally. The access conscious state serves as a part of a causal chain. The normal mental processes characterized by the cognitive sciences, memory, attention and the like, can be given a functional explanation in access conscious terms. So consciousness as access consciousness is readily available to a reductive explanation, hence the “easy” problem of consciousness.

¹⁴ It should be noted that characterized in this way, this form of explanation is not mere functional explanation but it is multi-level and thus comes close to a mechanistic explanation. For a mechanistic explanation, however, a mere possibility of searching for mechanism is not enough. Mechanistic explanations work as much with the realizers as with functional operations, and mechanisms thus directly affect descriptions at the functional level.

¹⁵ The terminology that I use is not the one Chalmers himself uses. Instead of phenomenal consciousness, Chalmers uses the term “consciousness”, and he uses the term “awareness” instead of access or functional consciousness.

Unlike access consciousness, phenomenal consciousness notoriously cannot be equally easily rendered in terms of functional causation. Many theorists have challenged functionalism as a theory of consciousness for this reason. Phenomenal consciousness or qualia, the subjective side of the experience has been taken to be something that cannot be given a functional role. The adversaries of functionalism have argued on basis of the nature of the scientific method that the subjectivity of experience eludes scientific method. Adversaries have also attempted to formulate various kinds of thought experiments to show how changes are possible in the subjective experience or qualia without a change at the functional level.

According to Thomas Nagel, functionalism and scientific method in general aim at objectivity, to give explanations from the third-person perspective (Nagel 1974 & 1986). But consciousness is essentially subjective, always in the first-person form. And because subjective and objective are each other's opposites, scientific method, when aiming at objectivity, drifts, in fact, farther and farther away from its object, the subjective experience of the first-person perspective. Thus it is impossible for such an objective method to conceptually capture subjective experience, let alone to give a reductive explanation of it.

Anti-reductionist philosophers have formulated several thought experiments to show why qualia cannot submit to a functional explanation, of which the most important also for Chalmers are the thought experiments of inverted and absent qualia, and the "knowledge argument". It is maintained that perfectly identical functional descriptions of their mental states can be given for two individuals whose visually phenomenal experiences are inverted. Descriptions of phenomenal experience are conceptually distinct from functional descriptions. Similarly for the thought experiment of absent qualia, but this time the phenomenal experience is suggested to be totally absent in the other individual. Both individuals are imagined to be perfect functional or even physical equivalents. There is allegedly no conceptual incoherence in such a scenario. Functionally or physically identical duplicates of persons without phenomenal experience are called "zombies". There is nothing it is like to be a zombie. Given the requirements of the reductive explanations above, there is no genuine global logical supervenience of qualia on the functional states in these cases because the phenomenal side can vary or be absent with no change in functional level. The upshot is that, if we accept these as genuine possibilities, the qualia, what-it-is-likeness or phenomenal consciousness cannot be given a functionally reductive explanation.

The famous "knowledge argument" by Frank Jackson (1982) introduces neuroscientist Mary who is a specialist in the neuroscience of color. It is presupposed in the thought experiment that she possesses complete scientific knowledge related to seeing colors. Interestingly, Mary is brought up and works in an environment that is on purpose held entirely colorless; her whole world has always been black and white, even though there are no deficiencies in her perceptual abilities. One day she is allowed leave her black and white prison and enter a normally colored environment. Jackson asks whether she

learns a *new fact* about colors at the moment when she steps into the colored world, namely, what it is like to see colors from the first-person perspective. If science could tell us what it is like to see colors from the first-person perspective, Mary would have possessed that knowledge, given her complete scientific knowledge about color vision. We have a strong intuition that Mary would learn something new, so this thought experiment, too, seems to suggest a conceptual separation between the scientific and phenomenal levels.

These thought experiments have been seen to constitute the most compelling challenge to the reductivist account. They are taken to show how the concepts of naturalists lack the required link to the level at which phenomenal consciousness resides. No co-variance and dependence is required between concepts, therefore global logical supervenience does not prevail. Accordingly, there can be no reductive explanation of phenomenal consciousness. Chalmers concludes that the problem of phenomenal consciousness is the “hard” problem of consciousness.

The easy and hard problems of consciousness are held to concern two main aspects of consciousness, those of the phenomenal and those of the functional aspects of consciousness. I concur with, e.g., Revonsuo (2006) that all the other concepts of consciousness are subordinate to or even subspecies of the phenomenal consciousness. There might thus be no easy problems of consciousness. I would agree with Chalmers that different forms of consciousness constitute different problems. However, I think explanations of consciousness do not have to be reductive, but a more systemic, multi-level model is needed. The issue has been raised by many that perhaps the hard problem of consciousness is not a *metaphysical* problem at all but rather merely a conceptual one. The problem is presented completely in the form of a good old *a priori* conceptual analysis, and, as was argued in section 2.1, many find it suspect that such an analysis allows conclusions to be drawn about the nature of worldly phenomena, but instead only as the consequences of our own intuitions. Chalmers has been sensitive to this problem, and has argued in depth that a certain kind of conceivability entails a metaphysical possibility (Chalmers 2002). My preferred answer to this line of thought is that the whole enterprise of making sense of the nature of concepts by starting from an analysis of the conceptual intuitions that we have is misled. It presupposes that the concepts and intuitions that we have are already correct in some relevant sense, and that they are not capable of development in the future. As Wilkes (1987) has argued, when conducting conceptual analysis, one must be aware of many background presuppositions and restrictions related to it. The possibility to develop is something that must be taken into account in the case of consciousness. We have no guarantee regarding in which direction the concept of consciousness will be developing in the future in the study of consciousness. I think that it is probable that when we or our successors finally get to grips with the hard problem, we will see the concept of consciousness differently, and it will not be possible to draw the same conclusions about the possibility of zombies (Van Gulick 2003).

The kind of conceptual development required is compatible with the co-evolution view of the mind-body problem. Concepts at the phenomenal level and neural and cognitive discourses affect each other by imposing bottom-up and top-down restrictions on the use of the concepts at the phenomenal level. Such restrictions may well prevent us from drawing such conclusions as are familiar from the zombie thought experiments and others of the same kind. This view is compatible with more recent views about the nature of explanation in philosophy of science, such as the non-reductive mechanistic explanations (Bechtel & Richardson 1993; Bechtel 2008; Craver 2005 & 2007; Revonsuo 2006). We might eventually be able to see how consciousness arises from the mechanisms that underlie it¹⁶. In the view of Revonsuo (2006), the problem of consciousness would profit from recent advances in the philosophy of science. The kind of explanation that we should opt for is not a reductive explanation either in the classical sense or in the looser sense. We could, in fact, admit the ontological irreducibility of consciousness and still hope for an explanation of consciousness. Mechanistic explanations purport to show what constituents contribute to consciousness and how. The model would be defective in that it would not apply universally; metaphysical possibility of zombies could not perhaps be completely ruled out, but this does not mean that the explanation fails. No prediction or identity is required; it is important to show how the given phenomenon works in light of the knowledge about its parts.

Furthermore, the nature of conceptual analysis depends on what the nature of concepts is taken to be. This will be taken up in section 3.4, and it will be argued that perhaps conceptual analysis should not proceed in the form of discovering counter-examples but instead it may be more instructive and contribute to our understanding to concentrate on typical cases. The consequences of this point for the problem of consciousness are obvious enough: thought experiments about zombies may be fascinating and intriguing but not perhaps to be taken so seriously when it comes to explaining or understanding consciousness.

Yet more, the adequacy of the concept of supervenience, on which the reductivist explanatory frame relies, has been called into question (Kim 1998; Horgan 1993; Lehar 2003). However, it has been claimed that an appeal to supervenience cannot guarantee materialism in principle, since supervenience is a negative principle that concerns only how mind and body are *not* related, not how they are related (Kim 1998; Lehar 2003). According to Lehar, "in fact the only epistemology that is consistent with the modern materialistic world view is an identity theory" (Lehar 2003, 379), by which Lehar means type

¹⁶ See the interesting visions presented by Revonsuo (2006) of just such a model. Revonsuo insists that the model has to be a first-person model, something to be enjoyed from the first-person perspective. This might work by loading the relevant data about the mechanisms of consciousness into a computer that would modulate one's brain to conform to that model. Then one would be able to enjoy the relevant aspects of consciousness from one's own first-person perspective, and thus see why some sort of constitutive mechanism gives rise to a certain kind of phenomenal consciousness.

identity. Given the familiar problems of identity theory, we seem to be faced with a dilemma, the two horns of which seem equally unappealing.

Another form of identity theory, *token identity*, is often held to be too loose about the relation between mind and body, even fitting autonomist positions. Moreover, it is not certain whether token identity is a coherent view: the mental and bodily types of token have distinct modal properties (e.g. mental tokens are essentially mental while physical are essentially physical) while all the properties should be the same for identical things. A related modal problem is one I like to call “modal multiple realization” (cf. Baker 1995). Suppose I eat a plate of oat-meal in the morning. As a result of metabolic processes during the afternoon, some of the proteins in the oat-meal become parts of neurons in the superior temporal sulcus of my temporal cortex. These neurons constitute the mechanisms necessary for my mental states when I later go bird-watching. If I decided to have scrambled eggs instead of oat-meal, and as a result of this, some (though qualitatively identical) molecules of the eggs end up building up some neurons of my temporal cortex, while my day would otherwise follow the same route. Now, a token identity theorist should insist, contrary to intuition, that the bird-watching mental states in these two scenarios would not be the same. If the modal properties differ, it means a violation of Leibniz’s Law (identical things share the same properties), by means of which identity is usually explicated.

Although supervenience, or identity, is unable to meet the metaphysical demands imposed on it, there may be some other positions available. Mechanistic explanations usually deal only with explanatory, not metaphysical issues, notwithstanding that in the core of scientific theories there arguably are ontological commitments that guide and restrict theory formation. Good candidates for metaphysical counterparts to mechanistic explanations might be *material realization* or *material constitution*. Whereas supervenience is a negative relation, realization and constitution are *positive* relations. Constitution and realization also retain relevant aspects of supervenience: the properties of the constituted things seem to supervene on those of the constituting thing. Realization also should be distinguished from supervenience in like manner (Beckerman 1992).

The difference between the two is that realization, like supervenience, is a relation between properties at different levels, while constitution is usually seen as a relation between things (Baker 2000). Realizations seem to better correspond to a top-down perspective, like functionalism. It seems even to require some sort of pre-existence of the abstract functional systems that become realized in different ways. Constitution, in turn, is more delicate, and requires a multi-level system in which there are different demands for existence at different levels. It is an *asymmetric, contingent* relation between two material things¹⁷, like a statue and the lump of clay constituting it. Asymmetric, as the

¹⁷ individual mental states may be taken as individual entities, so constitution may apply to token mental states and token brain states (Baker, 2000, p. 35, n. 3). Baker, however, does not elaborate this idea any further.

lump of clay can constitute the statue, not the reverse. Contingent, since there is no necessary relation between the two things: any lump of clay may come to constitute the same statue. Hence, the relation is non-reductive. Yet, there are not two things but one: the lump of clay *is* (in the sense constitution) the statue for a given period of spatiotemporal coincidence. The persistence conditions for the lump of clay and the statue differ: the lump of clay may exist before constituting the statue. The statue, in turn, allows replacement of its parts with pieces of other lumps of clay, e.g. when repaired. Moreover, when the lump of clay constitutes a statue, it acquires *new causal powers* in addition to its original causal powers, e.g. when it is sold, more is paid for it. The new causal powers disappear with the disappearance of the constitution relation. (Baker 2000.)

Constitution is not the mere (mereological) composition of parts. *The inner organization* of constituents matters. A lump of clay constitutes a statue only in a certain configuration, while composition can take several forms. Moreover, composition requires a certain set of components the identity of which its own identity relies on. The constitution of a thing can allow changes in its components as long as the relevant organization is retained. (*ibid.*) Constitution also allows this along a temporal dimension, and thus it is compatible with the kind of “modal” multiple realizability that token identity has problems with. It also allows other distinct modal properties for the constituting and the constitutive things: a statue is essentially a statue while the lump of clay constituting it is not. This is another advantage when compared to identity.

Constitution fits mechanistic explanations in the following respects. Both are based on the non-fundamentalist existence of things at multiple levels. Both are based on inner organization, and constitution allows articulation of the systematics of the relation between different the levels due to this, and in this way is more of the type-type than token-type. Mechanistic explanations, however, are based more on the workings of *some* of the constituents. I suggest that mechanisms could be seen as forming their own sort of constitution. The workings of the relevant parts constitute a phenomenon. Thus, the state of the entire brain does not necessarily constitute a phenomenally conscious mental state but only the workings of the relevant parts¹⁸. This could be viewed as “mechanistic constitution”. Mechanists themselves abstain from metaphysical considerations, or at least from the attribution of certain metaphysical labels (Bechtel 2008; Craver 2007; Revonsuo 2006), with the exception of Thagard (2008) who calls the metaphysical view of the mechanistic tradition *multilevel mechanistic materialism*.

These metaphysical considerations and explanatory issues in the recent philosophy of science accord well with the overall tenets of the philosophical

Although Baker’s view is based on thing-constitution, there are versions that view the relation as one between a phenomenon and constituting material, such as water and H₂O (Johnston 1997). Material-constitution is the more interesting one here, and Baker’s descriptions should apply to it *mutatis mutandis*. In the end, it is not clear whether this distinction makes any difference, since things can be constituted by many other things, like a man from his many organs, the organs from many molecules etc.

¹⁸ For the individuation of the parts deemed relevant, see (Craver 2007).

naturalism and mechanistic explanations. Semantic monism is guaranteed by multilevel integration in single theories rather than separateness of theories at different levels. The idea of co-evolution can be made more sensible with the idea of top-down and bottom-up constraints across different levels rather than reduction that requires the idea of identity or homomorphism (Craver 2005). It is in this framework that also Revonsuo (2006) envisions the possibility of future explanation of phenomenal consciousness.

2.4 Empirical approaches to consciousness

On the basis of the discussions in the previous section two broad approaches to consciousness can be distinguished: naturalism and anti-naturalism, depicted in “the consciousness tree” in Figure 1 (cf. Carruthers 1998a & 2000).

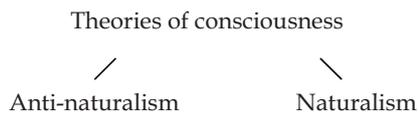


FIGURE 1 The consciousness tree

The present study aims to supplement conceptual analysis by taking into account the bottom-up restrictions that the empirical sciences impose on our concept of phenomenal consciousness. There have been numerous attempts in the neurosciences to reveal the neural basis of consciousness, to formulate a *neural* theory of consciousness. Therefore, they are the obvious place to look for such bottom-up restrictions. In the following, I will briefly review some of the relevant aspects of these theories of consciousness.

In many studies of the neural theories of consciousness, one or two forms of consciousness are taken as the starting point. The objective of such research is to reveal the neural processes that *correlate* with the respective form of consciousness (NCC, neural correlate of consciousness). The basic problem with this approach is that it cannot lead to any explanation of consciousness, neither phenomenal nor access consciousness. This is because correlation is not an explanatory relation. Neural correlates cannot explain consciousness, but they partly contribute to the project of explanation by providing a starting point and important restrictions for subsequent research and for theory formation. Revonsuo (2006) proposes that we should instead look for *constitutive mechanisms of consciousness* (CMC).

One of the most celebrated attempts to explain consciousness in neurobiological terms is the γ -frequency hypothesis coined by Francis Crick and Christof Koch (1990). The hypothesis is that consciousness is associated with the neural electrical activity that is observed, e.g., in EEG as activity in the

frequency of 30–100 Hz. EEG is a temporally accurate method for brain imaging, and a useful one for studying different frequencies of electrical activation in the brain. The γ -frequency was first discovered in the olfactory modality, but lately it has been discovered that the 30–70 Hz activity is widespread to all the sensory modalities (Crick 1994). It has been observed to correlate with conscious processing (e.g., Melloni et al. 2007). In relaxed wakefulness, the most conspicuous activity is α -wave activity ranging from 8 to 10 Hz. When something suddenly captures one's attention and requires more processing, the γ -frequency becomes the most prominent.

Nowadays, the γ -frequency is often taken to be the most reliable neurophysiological index of consciousness, although it seems that other brain rhythms such as α - and θ -oscillations are related to conscious processing (Palva & Palva 2007). In spite of being such an obvious index of consciousness, many have contested whether it even is a physical correlate of *consciousness* but rather a good candidate for solving the binding problem. The binding problem is the problem of how it is possible that the processing of distinct specialized systems, those processing, e.g., color and form in the brain is bound to a unitary perception. Brain processes input at several levels, and there is binding at at least three levels. At each level, there seem to be one or more distinct specialized processors. Let us take visual perception as an example. The first level of binding in visual perception could be the binding of several distinct dots leading to perception of a continuous line. Binding is internal to a specific level of a single sensory modality. This kind of binding is often taken to be hard-wired and not a result of learning.

The second type of binding is internal to one sensory modality but it binds the different features of that sensory modality into a unitary whole. In the case of visual perception this kind of binding is implemented by the specialized processors of color, form, movement etc. Binding at this level of may be the result of learning, as suggested by reports of blind people whose sight has been restored in adulthood.

Crick (1994) proposes this kind of binding to be the result of *correlative firing*. In *correlative firing*, the contents processed by groups of neurons that fire simultaneously at a certain frequency, such as γ -frequency, are bound together. Crick's examples are a visual sensation of a blue circle and a yellow square. It is well known that color and the form are processed in different areas, V4 and V3, respectively. For the forms and colors not to be mixed up as a yellow circle and a blue square or in any other way, Crick proposes that the neurons processing the color yellow and neurons processing the form square fire simultaneously, after which the neurons processing the color blue and neurons processing the form round in turn fire correlatively – and after them yellow and square etc. So the yellow and square are bound together, as are blue and round. On the larger scale, correlative firing is responsible for the binding of all the firing and thus processing of different contents in the cortex. Also at the third level of binding, synchronous activation plays an important role. This refers to the binding of all of sensory modalities to a unitary perception.

Another problem suggested for the γ -frequency hypothesis is that it is a correlate of access rather than phenomenal consciousness. The task of subjects is to report awareness of objects, and therefore the question is whether inattentional and marginal phenomenality might be associated with something other than γ -frequency activation (cf. Palva & Palva 2007). Empirical evidence of binding not necessarily being associated with the γ -frequency is given, e.g., by Revonsuo et al. (1997). They observed that γ -frequency is related only to the initial binding; it then dissipates while the perception of coherent objects is still possible. It is for this reason that only transient, global γ -frequency activation is taken to correspond to consciousness (Melloni et al. 2007). It also seems to follow that γ -frequency is not necessary for access, and so *a fortiori* perhaps not for phenomenal consciousness. It seems that γ -synchrony is somehow important but it is not clear what its specific role really is, and many remain skeptical about its ability to explain consciousness or even the unity-aspect of it (Canales et al. 2007).

Another popular form of NCC, in addition to the idea of correlative firing across many brain areas and sometimes directly related to it, is *recurrent activity* in neural networks. Recurrent activation in networks may include thalamo-cortical connections (e.g., Crick & Koch 1990) or cortico-cortical (e.g., Edelman 1989) or both (e.g., Lamme & Roelfsema 2000). A problem with attempts to define the relevant brain areas is that no such areas seem to be necessarily or sufficiently involved, and areas may vary. One possibility, then, is to define global activation functionally independently of specific areas. Some networks are mentioned more often than others. One such network concerns ventral and fronto-parietal areas in addition to sensory areas (“the ventral stream hypothesis”, see Goodale & Milner 2004). Even such global activation may not, however, be sufficient for conscious awareness; e.g., Huettel et al. found this sort of activation in non-conscious processing of changes (Huettel et al. 2001). Interpretation of such results naturally depends on what concept of consciousness happens to be at stake. It may be that this sort of activation, while not sufficient for phenomenal consciousness, is sufficient for access consciousness. The activation sufficient for phenomenal consciousness is almost intractable insofar as no clear index of phenomenality is at present available.

It has been suggested that mere activation of V4 is sufficient for phenomenality (Zeki 2003), and if this is “functionally” the case, then we cannot tell whether there are isolated areas in the brain that are conscious, and with no access for the subject to report on them or be aware of them. It is frequently noted that many different kinds of neural correlates or mechanisms for phenomenal consciousness may exist, and thus many little phenomenal consciousnesses may be living in the brain some of which can be at any given time bound together. Unraveling this poses a major challenge for anaesthesia and consciousness studies (Stoerig 2005). This issue is not unrelated to the question of consciousness being a dimension rather than a categorical distinction between conscious and non-conscious states (e.g., James 1890; Dretske 1995).

Students of consciousness have regarded both the γ -frequency hypothesis and global activation of recurrent networks theory or their combination as good candidates for explaining binding and correlate of memory, short-term or sensory memory, in addition to access consciousness. Needless to say, these amount to no more than the starting point for a more fundamental explanation of how subjective experience can be related to this activity. If we take the strategy of CMC instead of NCC, the major change is that any systematization becomes unnecessary. Finding mechanisms for specific instances of conscious mentality may be possible piecemeal, while offering no general prescription for it. This does not mean that no such prescriptions could exist, and descriptions of mechanisms at a more general level may count as such prescriptions in particular instances.

In addition to all these various neural approaches to consciousness, consciousness can be approached empirically by means of the cognitive sciences. The most famous and the most salient theory for the present study is the *cognitive* (functional) theory of consciousness formulated by Bernard Baars (see, e.g., Baars 1988; 1994; 1996; 1997 & 2002). Cognitive theories form another branch in the consciousness tree.

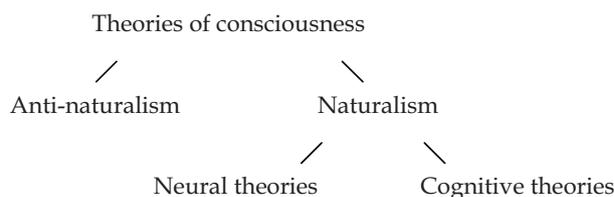


FIGURE 2 The consciousness tree, version 2

Baars suggests that the cognitive function of consciousness can be discovered by experimental cognitive science. Experimental studies require that the object of research (or an index of it) must somehow be treated as a variable. Consciousness, too, must be treatable as a variable. The variance of the variable must be able to be rendered on a measurement scale. The simplest form of the most basic one is the nominal scale, which gives the variable values 'yes' or 'no'. The problem with consciousness is that it is difficult to say when to give consciousness even these values. The choice of measurement scale for consciousness should not perhaps be that simple. Consciousness may admit of being measured on an ordinal, interval, or even ratio scale, although it is very difficult to give consciousness any precise value. This should be possible in principle, e.g. in reference to the distinct concepts of consciousness discerned above, and is an important challenge for the future science of consciousness (Ramsøy & Overgaard 2004). The major limitations of the theory of Baars stem from this limitation, but it offers a sensible starting point to evaluate the broad absence or presence of consciousness and the difference between these.

Baars calls the comparison between presence and absence of consciousness “contrastive analysis”. Originally a psycholinguist, Baars is vividly aware of the abundance of the non-conscious mentality. Psycholinguistic experiments have revealed an enormous amount of non-conscious processing going on when for example the syntax and semantics of a proposition are processed. When a part of a single word is seen, the part activates a number of possible candidates for words and constrains the determination of the unconscious contexts through which our conscious linguistic processing is facilitated and made possible. The non-conscious processing underlying conscious mentality consists largely of the same kind of processing that is present in the conscious processing: e.g., recognition, perception, and planning. In other words, there seem to be non-conscious counterparts to many forms of conscious mentality. Non-conscious processing, however, differs in many important ways from conscious processing. By comparing conscious processing to its unconscious counterparts, we can empirically form a theory of conscious mentality¹⁹. This comparison constitutes the method that Baars calls contrastive analysis of consciousness²⁰, and by this method he has formulated his theory of consciousness.

The core of the theory is the following. Whereas consciousness is characterized by limited capacity but vast access, non-conscious mentality is characterized by limited access and vast capacity. This means that there can be only little content in consciousness at any one moment. But the content of consciousness is largely distributed: there is a “global availability” to the contents of consciousness. Behind the claim is the commitment to the view that the mind (or brain) is a collection of parallel distributed processors. There is a consensus among all the “mind sciences” that mind consists of several quite independent systems that specialize in processing only one type of input, as was seen above in the case of visual awareness. According to Baars, there is global availability in the sense that all of these specialized processors have *access* to the contents of consciousness. He grounds this claim in experimental evidence; it has been shown, e.g., that people can submit almost any part of the brain to conscious control provided that they get conscious feedback. (Baars 1988 & 1997).

Non-conscious processing, in turn, is very fast and effective and, compared to conscious mentality, it makes fewer mistakes. In fact, when mistakes are made, or something unexpected happens, non-conscious processing turns into conscious processing. Consciousness is needed most often when something new is confronted and more resources are needed to process it. Here, Baars admits his debt to the famous AI theorists Newell and Simon

¹⁹ Although Baars holds that his theory is a “thoroughly empirical theory”, he, nonetheless, proceeds to make a hypothesis to account for the wealth of relevant data. This way he actually attempts an abductive inference to the best explanation. His theory is thus not devoid of any theorizing, although it is thoroughly based on experimental data.

²⁰ Baars claims to follow here the example of William James’s contrastive analysis (Baars 1997). According to Baars, James made the crucial mistake of holding the unconscious mentality to be purely mechanistic like reflexes. He did not clearly enough see the resemblances between the conscious and unconscious mentality.

who demonstrated that the problem-solving potential of specialized processors working at a problem accumulates exponentially when the number of specialized processors is increased (Baars 1988). The efficiency of problem solving is thus optimal and at the same time the system works very economically when routine tasks are performed by just a few specialized processors. The unconscious processors are thus very independent: the working of some parallel processors does not interfere with the processing of the others. In the problem-solving case this is an advantage, however, because then the most efficient resources are found faster.

Conscious processing results ultimately in automaticity of processing: it turns into a non-conscious routine. This is well exemplified in the phenomenon of sensory or semantic satiation. We rapidly get used to olfactory and haptic sensations such as those of our clothes, different temperatures and the smells of surroundings, even quite disturbing ones. Sensory satiation works in all of the sensory modalities. However, the visual system protects itself from satiation by micro-saccadic movements. If movements of the image constituting the visual field are adapted to follow eye-movements, the visual sensation becomes satiated: it wears off. The visual system seems to have some privileged place in consciousness. If there is any rivalry between the contents of the different sensory modalities, the content of the visual modality seems to be more easily selected.

Baars calls his theory the “global workspace theory of consciousness” (GWS). The workspace is a working memory system in which the contents are available to all the specialized processors – thus the workspace is global. The limited capacity of conscious processing stems partly from the limitations of working memory. Another source of limitation is selective attention whose function in the theory is to select the contents of the working memory. The cognitive functions of memory, attention and consciousness are distinct; memory serves to retain the contents, attention selects the contents and the function of consciousness is to guarantee its vast access.

Whereas selective attention chooses one of the contents of the specialized processors to provide the content for the global workspace of the working memory, the non-conscious mentality of the other specialized processors form the *context* that determines the contents of consciousness. According to Baars, such contexts exist at many hierarchical levels. Deeper contexts determine the contexts above them. For Baars, automaticity of processing means that it becomes one of the contexts that determine further contents of consciousness. Contexts thus include the previous experiences and knowledge structures of the subject, and the action goals of the subject are formed on their basis. Contexts that become problematic can come back to be revised in consciousness.

According to Baars, the deepest and the most general context is the *self*. The self consists in our past experiences and goals, our relations to other people and so on. Thus it corresponds well to what many theorists have called the narrative or extended self (e.g. Damasio 1999). The narrative or extended self is a high-level construct that seems to require episodic memory to become formed. The concept of self in Baars, however, is larger in that it also counts the

bodily self as one important aspect of the self. What all these have in common is that they mold and modify our experiences and make consciousness of other things possible. A sudden disturbance in any of these, a sudden obstacle to an important goal or a serious physical injury requires one to pay attention to these deepest contexts behind our consciousness, thereby making them the contents of consciousness.

The GWS theory is closely linked, also by Baars, to neural theories, especially to the γ -frequency hypothesis and recurrent processing. Other interesting suggestions have been put forward regarding the neural mechanisms for global availability, such as specific workspace neurons with long axons connecting local networks with the prefrontal and parietal areas (Dehaene et al. 2003; Dehaene et al. 2001). Here, the GWS model links together different levels of description to one multilevel mechanistic model, and it is thus a good example of theory formation as described in the “mechanistic tradition”.

The standard objection leveled against the GWS theory is that in many cases of unconscious mentality there is a sort of access that is not limited to routes that are intramodular or specific to a certain specialized processor. It is unclear, however, whether these cases can exhibit global access with the neural properties of the γ -frequencies associated with it.

Another important objection is that GWS is rather a theory of access rather than phenomenal consciousness (e.g., Chalmers 1997). I think that this is a misunderstanding, although it is fair to say that Baars himself does not take these different forms of consciousness into account. Others, however, have noted that when understood correctly as global *availability*, it becomes clear that access consciousness could be seen as an instance of real access, whereas phenomenal consciousness can be seen as that part of consciousness that is available for access but not necessarily accessed (Prinz 2000b; 2003; 2004). The neural theory of consciousness of Jesse Prinz takes this form. It is interesting to note that, in the view of Prinz, the accessed contents are not necessarily conscious at all but only their imagistic accompaniments. In the same vein, Baars describes imagistic phenomenal contents as constituting the key to the accessing most abstract contents (Baars 1988, 244). It is not clear to me whether Baars sees that these abstract contents themselves as conscious or not, although the GWS theory certainly allows such an interpretation.

Baars’ theory is in my opinion a good example why something more than conceptual analysis is needed. We need empirical studies so that conceptual analyses would not merely reflect our own presuppositions and to get to know about things and aspects to which we have no direct access. This way we can enrich our psychological concepts. It seems to me that it is through empirical studies that most progress can be made concerning our concept of consciousness. The theory of Baars is a good example of the value of the empirical data, and it has contributed to philosophical analyses of consciousness. Many theories of consciousness, more or less philosophical, have adopted the general approach (e.g., Dennett 2001; Chalmers 1995). My more detailed suggestion in the end of this study for a distinct, more mundane form

of consciousness, “mentalized consciousness” also leans largely on the idea of availability and contexts.

3 THE REPRESENTATIONAL THEORY OF CONSCIOUSNESS

Empirical approaches to consciousness conceal another important distinction. This distinction is related to the question of the semantic or representational properties of consciousness.

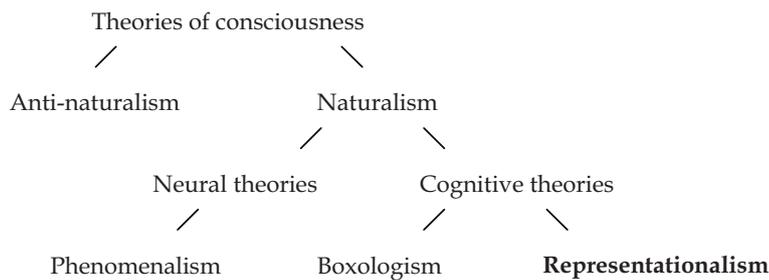


FIGURE 3 The consciousness tree, version 3

During the last few decades, the so-called representational theory of mind has probably been the most widely held view amongst mind sciences. The central tenet of the theory is, naturally, that mind consists of internal representations of the external world. Psychological processes consist of the processing of these representations. Most branches of contemporary psychology, and especially traditional cognitive science, are based on the representational theory of mind. Cognitive theories of consciousness also include a purely cognitivist or functionalist approach, one referred to, especially by neurologists, as “boxologism” takes consciousness not necessarily referring to representational properties of consciousness^{21,22}. Merely neural theories are usually

²¹ The cognitive theory of Baars is mainly neutral with respect to this distinction, but different versions of the general global availability approach can fall within either the domain of representational or boxologist theories. One example of the former is in (Tye 1995).

²² Some recent radical externalist theories of consciousness in the cognitive sciences also are anti-representational in that they situate consciousness in the environment rather

“phenomenalist”²³ in that they try to explain phenomenal states by reference to neural states, and not see possible semantic or even functional properties of phenomenal states as necessary or relevant.

As far as phenomenality is concerned, traditional analytic philosophy has seen this as the non-representational part of the mind. Behind this view is the old conviction that only linguistic, conceptual contents can truly be representational or intentional, since only they can have criteria for truth and falsity. Recent thought experiments concerning the conceivability of zombies and inverted spectra have only strengthened the view that phenomenality does not contribute to intentional properties.

Recently, however, this traditional conception has been breaking. Representational theories of phenomenal consciousness have also tried to analyze phenomenal consciousness through the concept of representation. It is this theory of consciousness, of which the higher-order theories are a subspecies, and a version of which will ultimately be applied in formulating the theory of mental self-knowledge, which is the main topic in this chapter. In the following, I present the main ideas of the representational theories of phenomenal consciousness, referred to in brief as *representationalism*²⁴ or *intentionalism*. I then show how it can cope with the traditional challenges of the representational theory of mind. The rest of the chapter deals with topics that are relevant for the topics of the subsequent chapters. Sections 3.3 and 3.4 deal specifically with the nature of representational *properties* and *contents*. Finally, I turn to some new problems of the representational analysis of phenomenal consciousness that seem to plague the representational theory. In order to cope with some of these problems, representationalism has been developed to cover two levels, those of the first-order and higher-order (representational) theories of consciousness.

that inside the subject (e.g., O’Regan & Noë 2001). The idea is that all the relevant information about the world is in the world itself, and that the processing of mental representations by the subject is not required. The subject can use that information directly to cope in the world. These ecological theories are difficult to situate in the consciousness tree but perhaps deserve ramification in the Naturalism branch.

²³ ‘Phenomenalism’ here should not be confused to the classic ontological position with the same name famously held by, e.g., Berkeley.

²⁴ Representationalism is also the label of the epistemological position of indirect realism. Modern representationalism is, in fact, not an epistemological position. If it were, it would rather be a form of direct realism. Thus, it does not suffer from the classical problems of epistemological representationalism, such as the infinite regress of ever diminishing homunculi: modern representationalists have pointed out that representations are what we are conscious with, not what we are conscious of (Dretske 1995, 100). This way, no inner perceivers need be postulated.

Some, like Lehar (2004), see modern representationalism as corrupting the traditional sense by their using of the term and that this leads to many problems.

3.1 Mind and intentionality

The spirit of the contemporary representational theories of consciousness is captured in the two clauses of the Representational Thesis formulated by Fred Dretske (Dretske 1995, xiii):

- (1) All mental facts are representational facts
- (2) All representational facts are facts about informational functions

With the clause (1), the thesis purports to cover phenomenality. For a full appreciation of the representational thesis and especially of (2), a brief outline of what Dretske means by representation is needed (Dretske 1981; 1986; 1988 & 1995). Most importantly, the mark of representational systems is that they are capable of *misrepresentation*. As an informational theory, Dretske's theory leans on the concept of indication. Indication is the natural property of different kinds of objects of carrying information about something else. A doorbell ringing is the doorbell's property of indicating or carrying information about the doorbell's button being depressed. Similarly, the doorbell's ringing indicates (or carries the information that) the doorbell is not being left undepressed. The latter case shows that the indicational or informational relation is not necessarily a causal one, but can be analytic or nomological. In addition to man-made indicators, the natural world is furnished with indicational relations between the elements in it. Traces of hares in snow indicate recent presence of hares; annual rings of trees indicate age of trees etc. A signal such as ring of the doorbell or traces of hares' footprints on the snow can indicate (or carry information about) many things, indeed an infinite number of things as long as they have a reliable connection with it. By reliable it is meant necessary; nothing less suffices for an informational relation. As a result, an indication relation does not or, better, cannot fail to indicate something. If it were possible to fail, there would be no guarantee of information having been carried. An indication relation is all or nothing; there can be no misindication.

Dretske emphasizes that representational relations are indication relations that have the *function* of indicating something. Only functions have the possibility to fail. An indicational relation is promoted to a representational relation once the indication relation acquires a certain function of signalling the presence of something specific connected to its source. Annual rings of trees indicate the age of trees but they do not represent them unless they have a function to indicate them. A lumberjack, who wants to make notes of the ages of the trees he has cut down, can assign this function to the rings by his note-making. It is possible that during some year the tree has not grown because of abnormal weather conditions. Then the number of rings fails to mean the age of the tree as wanted by the lumberjack. The action of the lumberjack thus makes misrepresentation possible. The intentionality of the annual rings, however, is only *derived* from the intentionality of the mental states of the lumberjack. The mental states of the lumberjack are not derived from anywhere but are an

example of natural original intentionality. Explaining natural functions or natural intentionality is much more difficult and interesting.

Original or natural intentionality is often viewed as a property of biological organisms. A popular example is the frog's ability to detect flies in its environment. The frog has certain neural structures that are called 'bug detectors', structures that are sensitive to the presence of small black moving objects in the environment. The neural structure is an example of a natural indicator; it indicates the presence of bug-like objects in the environment at a certain range. Moreover, the bug detector is coupled to a certain behavioral response, one of catching the indicated objects with its tongue. In the normal environment the frog can be said to represent flies, and the behavioral response should benefit the frog. In a laboratory, the bug detectors fail to produce the same effect: they misrepresent all small moving objects as flies (or something edible).

The concepts of truthfulness or criteria for accuracy are essentially normative unlike the traditional concepts in the natural sciences. In analytic philosophy, normativity and the criteria for meaningfulness are seen as possible only through social conventions or contexts of social use. In a sense, however, a frog catching flies is a context of use itself. The representations of the frog are recruited for use, and use creates functions. This kind of normativity (criteria for success) seems to be one that fits the natural sciences like that of evolutionary success or functionalism, which this kind of success is often identified with. This kind of analysis of representation seems possible in terms of natural science.

The leap to this kind of normativity or the question of how to acquire a representational function is famously problematic, though. The standard way of making sense of this is by the teleo-functional approach in which it is some kind of success that allows the move from indication relation to representation. For example, one can appeal to success in evolution (Millikan 1984). This view, also often ascribed to Dretske, however, suffers from conceptual problems such as the one posed by the Swampman thought experiment of Davidson (1987). The Swampman, originally intended as a counter-argument to theories of reference that appeal to causal history, is an imaginary person who is supposed to be a perfect physical duplicate of a person like the philosopher and semanticist Donald Davidson himself with the exception that he lacks the personal and evolutionary history of Professor Davidson, since he came into existence as a result of a stroke of lightning in a swamp, where the electrical discharge is supposed to have rearranges a bunch of molecules to correspond perfectly to the molecular structure of Davidson himself. Whether we take this "Swamp-Don" to have intentional states or not, has direct consequences for the validity of theories based on causal or evolutionary histories (for a discussion of applicability of this thought experiment to Dretske's theory, see Adams & Dietrich 2004). The counter-argument does not, however, concern the teleo-functionalist approach as a whole. There are ways of making sense of natural functions or representational content that do not appeal to historical factors but only to success. Teleo-pragmatic views that think of content only in terms of

success in action are not affected by the Swampman's lack of any sort of history. In this view, action and success, the functional properties of the Don and Swamp-Don are as indistinguishable as their structural properties.

Galen Strawson has contested any action-based theory of mental states by the famous "Weather Watchers" thought experiment (Strawson 1994 ch. 9). Weather Watchers are a race of organisms as described by Strawson that have lost their ability to move in the process of phylogenesis. Nevertheless, they have retained their perceptual and cognitive capacities. They could be described as a sort of conscious fungi. What they can do with these capacities is no more than observe passively their surroundings, mostly changes in the weather. If we accept the possibility of the existence of such creatures, their mental states cannot depend on action, not even dispositions to act. If they can have minds, then action cannot be essential to mentality in general.

Fortunately, the general concept of success includes more than success in action. There might be creatures that benefit from not acting at all. More importantly, there are aspects of mind whose success is bound to more intramental properties. One is *learning*; the capacity to learn may be assessed by success and thus is the source of criteria of accuracy for representational content. This is, in fact, what Dretske himself relies upon, though he often gets lumped with those explaining function in terms of phylogenesis (Dretske 1988). Another candidate for source of criteria of accuracy could be *control* (Bechtel 2008). Here, control means control of external or internal factors, and both can be assessed by success.

Together these two theses (or their combination) entail that instances of consciousness may be *causally* dependent on evolution and action but not *ontologically* dependent (Revonsuo 2006). It is true that our minds would not have their mental properties or exist in the first place without the evolutionary history or successful action. However, other kinds of minds might exist whose causal history of existence is pure accident or being intentionally fabricated.

Originally, Dretske formulated his views to explain the representationality of conceptual mental states, such as beliefs. Later on, he applied this framework to phenomenal consciousness, the dimension of mind that was previously held to be unintentional. He insisted that both the conceptual and non-conceptual aspects of consciousness are intentional, and intentionality is a representational/informational function that permits a naturalistically acceptable explanation. Properties of mental states that are traditionally seen as phenomenal and non-conceptual (qualia) are (identical with) properties of the represented objects. Dretske calls these properties "representational" properties of objects, reminiscent of secondary properties.²⁵

²⁵ According to Chalmers (2004): "Some representationalists, such as Dretske and Tye, occasionally put their view by saying that phenomenal properties are identical to certain represented external properties, such as physical redness. As I like to put things, that would be a category mistake: phenomenal properties are by definition properties of subjects or of mental states, and physical redness is not (or need not be). I think that this is simply a terminological difference, however. For example, Dretske defines phenomenal properties ("qualia") as the properties we are directly aware of in perception, and concludes these are properties such as colors. This is quite

This straightforward form of representationalism is sometimes differentiated from more modest analyses of mind in representational terms. Another proponent of this form of representationalism, Michael Tye formulates the difference like this:

“Representationalism is a thesis about the phenomenal character of experiences, about their immediate, subjective ‘feel’. At a minimum, the thesis is one of supervenience: necessarily, experiences that are alike in their representational contents are alike in their phenomenal character. So understood, the thesis is silent on the nature of phenomenal character. Strong or pure representationalism goes further. It aims to tell us what phenomenal character *is*.” (Tye 2000, 45)

While the strong representational theory identifies phenomenal properties of experiences with representational intentional properties of the content – the phenomenal qualities that make up our perception are the represented properties of the objects – the more moderate thesis makes the same suggestion with supervenience. This could be called weak representationalism²⁶.

Tye’s own view is a strong one and corresponds to that of Dretske with little differences in detail. He describes phenomenal contents as poised, abstract, non-conceptual intentional contents (“PANIC theory”). This adds that phenomenal contents are available for use (poised) as in the GWS theory.

The first and foremost reason for endorsing representationalism is the *diaphanousity* or *transparency* of experience.

“(TE) The primary introspectively accessible aspect of a phenomenological experience is its world-directed representational content.” (Kriegel, in press)²⁷

Every phenomenal aspect or quale cannot be detached from what it is about. If you try by introspection to confine the redness of a ripe tomato or the bitterness of a lemon to pure sensory quality ‘redness’ or ‘bitterness’, the experience resists this. The redness tends to remain as the redness *of* the tomato and the

compatible with the claim that phenomenal properties in my sense are representational properties, as long as one holds that one is directly aware of the represented property rather than the representational property. Once we make the relevant translation, I think that these representationalists’ most important claims can be put in the terms used here without loss.”

²⁶ Seager & Bourget (2007) call theories that take all aspects of mind as representations “pure representationalism”, while impure representationalism is the view that there are some non-representational properties of mind, such as manners of representation (cf. Chalmers 2004). Most representationalists are impure representationalists.

²⁷ The formulation is due to Kriegel (2008) in which the original formulation reads as follows: “(TE) The only introspectively accessible aspect of a phenomenological experience is its world-directed representational content.” I think it is a bit exaggerated to say that we cannot introspectively access any aspects of phenomenal experience other than representational content. In my view, it is possible to access at least the *modality* of the phenomenal experience (see section 5.1); one can tell whether one sees, hears or believes by introspection, and scrutinize the difference between different modalities. Also, some introspectively available temporal aspects of experience may differ from those of its contents. TE, however, is correct in that these are secondarily given aspects of the introspected state. Kriegel himself, to accommodate the transitivity principle, holds the odd view that consciousness is peripherally given in conscious states, but this givenness cannot be introspected.

bitterness that of the lemon. The detaching of the quale seems to be possible only at the conceptual level, that of *phenomenal* concepts. I want to emphasize that this property of the experience is fully compatible with the relational-representational view of phenomenal consciousness where the phenomenal aspects of consciousness are determined by their relations to things in the world and their relations to one another. The intrinsic properties that many wish to attribute to them are due to the physiological vehicles specific to these representations. This specificity, however, depends wholly on the physiological basis of the mind that it is a part of, and its relation to the environment.

Representationalism is, primarily, a reductionist strategy; it reduces phenomenal properties to more basic representational properties. Representational properties can be given a naturalistic explanation if one can provide a naturalistic theory of representation. The informational teleosemantics strategy is the choice of Dretske, and causal covariance strategy is adopted by Tye (1995 & 2000). Reduction is guaranteed by the identity of phenomenal and representational facts. A reduction by conceptual entailment is provided with at least a supervenience or identity relation between the concepts in question.²⁸

²⁸ Representationalism is reminiscent of another tradition that has little to do with reductionism. Also in the continental phenomenological tradition, starting from Brentano, intentionality has been seen as the defining feature of mind. Phenomenology and representationalism differ, however, in some aspects. Intentionality means slightly different things in each tradition (cf. Dretske 1995, 32-33). In the phenomenological tradition, intentionality is seen to include two elements according to an act-object model. Noesis is the act or the process of intentionality, and noema is the intentional object of the act. The phenomenological view is then a form of the act-object model of mind. The nature of noema is debated, however, mostly in regard to whether Husserl conceived it in an internalist or externalist manner (Zahavi 2004). Be that as it may, intentionality in the phenomenological tradition is seen as something worldly. "World" means for them something perspectival related to a *Dasein*, a body, or whatever, but, in my view, this muddles the debate about internalism and externalism. I follow Barry Smith (MS) in that this conception of world is not easily compatible with the concept of world in the natural sciences. In contrast, phenomenologists themselves take the life-world to be primordial and the world of the natural sciences to build on the life-world. This is of course true if we think of the relation of the subject, *Dasein* or body on one hand, and the world of the natural sciences on the other. Nevertheless, this kind of view cannot account for the world as independent of our existence as the natural sciences have revealed it to be. Therefore, the phenomenological tradition clearly gives, in addition to epistemological priority, also ontological priority to the life-world. They emphasize that the life-world is as existent as the world of natural sciences. I agree with Barry Smith that this view does not take in a sufficient manner into account the part-whole relation that the life-world and the world of the natural sciences have or should have to one another (*ibid.*). It seems quite incontestable that the world of the natural sciences should bear the ontological priority in being the whole world whose part the life-world, world as the environment of a living organism, is, as emphasized by Smith. Epistemologically, the order may be inverse, the appearing life-world serving as the basis of our knowledge of the reality behind appearances. The issue is perhaps not this simple, since at least at the later stages of development we hold the natural attitude that the world we perceive and theorize about is real. It takes a great cognitive effort to suspend this attitude and reflect on the manifesting practical life-world on which our theoretical world-view is built.

Unlike most phenomenologists (exceptions include Thomasson 2000 and Thomas 2003), contemporary representationalists do not conceive of the mind by an

As an addendum to the representational thesis, it is a popular strategy among representationalists to respond to the zombie and knowledge arguments by appealing to “the phenomenal concepts strategy” (e.g., Loar 1997; Papineau 2002; Tye 2000; Carruthers 2000). Phenomenal concepts are those concepts by which it is possible to characterize and categorize phenomenal properties of experiences in contrast to sensory concepts that apply to sensory properties of objects. In the common notation, “Red_p” is distinct from “red_s” in that the former pertains to redness of the experience while the latter pertains to the properties of experiences. The phenomenal concepts strategy maintains that it is the properties of phenomenal concepts that allow conceivability of zombies and new knowledge for Mary. These, however, do not have to carry over to the ontological level of phenomenal consciousness itself. By postulating a metaphysical identity, a reduction can be provided. The hard problem of consciousness can be conceived of as only a conceptual, not necessarily a metaphysical problem. If one accepts representationalism as one’s metaphysical theory of consciousness, one does not have to worry about conceptual problems implied by phenomenal concepts. The validity of representationalism lies mainly in the conviction that an identity relation exists between phenomenal and representational facts.

Representationalism does not, however, have to be dressed in the guise of reductionism. Representationalist analyses have now become popular even amongst staunch anti-reductionists like Chalmers and Jackson. Representationalism can be defended as a general analysis of phenomenal states, but it cannot be held, however, that the phenomenal states are reducible to physical properties like those of the physical objects constituting the content of the representation. One can also hold a version of representationalism that goes only halfway, maintaining that the properties of objects are perceiver-relative in a way that allows for inverted spectra, and thus allows a more traditional way of conceiving qualia (Shoemaker 1997). Anti-reductionist intentionalistic attempts also include some recent attempts to analyze consciousness in the phenomenological tradition (Thomasson 2008), and also representation in the mechanistic tradition (Bechtel 2008). It is in their spirit that my own view will follow, since they incorporate the first-person view, representation and the kind of explanatory strategy that I favor.

Some prominent anti-reductionist representationalists typically resist the phenomenal concepts strategy. I will return to this issue in the following section after a brief discussion of some semantic topics.

act-object model. Intentionality means for both traditions *aboutness*; every state of mind is purportedly about something other than itself. Contemporary representationalists, however, take this to be something in the world in the more traditional sense: intentionality is based on criteria of accuracy that are determined in part by the outer world. The phenomenological tradition mainly neglects this aspect and concentrates on the phenomenal “alterity” of intentionality: mental states are directed at something other than themselves.

3.2 Representational content: narrow or wide

It has become popular to analyze the concept of representation by distinguishing the *vehicle* from the *content* of representation. The vehicle (the brain) is always conceived of materialistically while the content need not bear any similarity to its vehicle (e.g., Dennett 1991). Some views distinguish the two completely and some hold them as interdependent. The representational vehicle is, in the case of (human) mind the brain, conceived, e.g., as a neural network. Content is seen as partly dependent on its relation to other representations or on its relation to the outside world (or both). The first option is an internalist (narrow content) view of content and the latter an externalist one (wide content).

The question of the wide and narrow content of representation is a question of how we should individuate the content of a representation. One good way to characterize the difference between narrow and wide contents is in regard to indexical thoughts (Carruthers 2000, 104-5). Suppose you are looking out of a window and see a blue tit perching on a branch; you then avert your gaze for a second and then look back. Meanwhile, the blue tit has flown away, but another happens to be perched in exactly the same spot. Adult blue tits look alike even to an experienced eye²⁹. You cannot tell the difference between the two blue tits, and probably will consider them as one and the same individual. The question is: does the *content of the phenomenal experience* differ in each of these instances? According to the narrow content view, the two experiences are indistinguishable and thus have the same content. On the wide content view, one is compelled to say that the contents of the two experiences must differ from each other since the object partly determines the content and each experience had a different object.

The most important reason for endorsing narrow content view is its suitability for psychological explanation. Suppose again that you would decide to slowly withdraw from the window in order not to disturb the little bird. The existence or number of birds does not, however, matter if we are interested in explaining your action at a psychological level. What is important is what *seems* to you from the first-person perspective. From the third-person perspective, narrow content is often conceived of holistically as determined by the role of the contents in the internal system of the perceiver. Even though this is quite natural, it is not necessary. Narrow contents can be atomistic, e.g. in the case of purely recognitional concepts that are only determined by the contents of one's occurrent perception.

The narrow content view is often contested by questioning whether it is a form of content at all since it is not interested in truth-conditions. As far as narrow content is concerned, the existence of and resemblance to an object is irrelevant. What has been individuated in the case of blue tit, is a totally

²⁹ In fact, there is a specific identifying feature that distinguishes the male blue tit but it can only be seen with a device that renders ultraviolet visible.

subjective mental image of a blue tit³⁰. There are at least two ways of seeing narrow contents of this kind as contents. Chalmers has suggested that narrow contents obtain truth-values by epistemic relations to other narrow contents of the subject (Chalmers 2002b). The critics of this kind of view think that this commits us to the view that we should be able to derive every other content from a single narrow content (Botterill & Carruthers 1999, ch. 6). I think this is an exaggeration. We can think that qualitative differences of consciousness (mental representations) are determined by differences and relations to other representations and qualities in order to have the ability to *discriminate* (cf. the discrimination theory of color in Arstila 2004). We have learned from physiology that differences in colors are due to the kind of cells that process the wavelength of light that is specific to a color. And the opposite colors are processed by cells more similar than those processing colors adjacent in the color circle. Red was separated from green when a mutation slightly changed the cone cell that processed green. There is also a necessary interdependence between different elements of color. Hue, saturation and brightness together constitute the color and they cannot be separated from each other. At a yet higher level it seems that the properties of colors depend on the entire mental sphere and the place they occupy in it; as Carruthers points out, red is felt as warm and blue as cold, and this can only be understood holistically, not atomistically (Carruthers 2000, 80–81). And due to this, a spectrum inversion is perhaps not possible in the case of one individual without a difference in function. If red is inverted to green, then perhaps the individual would notice the change and green would start feeling warmer. It seems fair to say that such interdependence and interdetermination is a fact in the representational system that mind consists of, but not that radical derivability should follow from this.

Other propositions in support of a narrow content view include response-dependent properties (Kriegel 2007; Prinz 2000a) and phenomenal properties (Siewert 1998). Revonsuo (2006) suggests interestingly that perhaps we should view mind as a sort of “container”. Whatever is contained in it amounts to a sort of internalist, yet non-representational content. My own leanings are toward the account of Siewert; in my view, phenomenal properties *can* be seen as criteria for accuracy, although this is not necessary. Introspection also requires this sort of content: a phenomenally manifested content but in which the emphasis is not on its truthfulness, but at most its potential truthfulness.

The proponents of the wide content view, in turn, emphasize that external factors at least partly determine the content of the representation. The lesson of semantic externalism (e.g., Putnam 1975) that has come to prevail in recent years is that facts and things in the world determine what our representations are representations of, regardless of their accuracy. Externalism, however,

³⁰ It is not even allowed to say, in the spirit of Searle (1983), that we have individuated “the blue tit that my experience is about” because to do so we must have metarepresentational concepts like “experience”. It is not necessary to have such concepts in order to have experiences with narrow contents. And even if one did allow those concepts it would be unnecessary and far too complicated to suppose that such a self-reference intrinsic to every experience exists. (Carruthers 2000, 105.)

seems inevitably to suffer from problems of misrepresentation and disjunctivism (see, e.g., Martin 2006): what is the common factor in two qualitatively identical experiences of blue tits of which one is veridical and the other not? Pure externalism does not seem to have the resources to explain this. Internalism has a natural way to distinguish between them: they have the same narrow content.

Lehar thinks that every attempt to endorse a sort of wide content is doomed to failure for the same reason as every attempt to explain consciousness naturalistically in any way other than by means of an identity theory, e.g., in terms of supervenience. Such a position, in Lehar's opinion, is committed to nothing less than Cartesian dualism because only representations or their objects are materialistically conceivable. Once you try to reconcile the two positions you have to postulate something whose existence is difficult to conceive, let alone demonstrate. This something is normally the content of consciousness or representation, not the vehicle of representation as distinguished by Dennett (1991). According to Lehar,

"[T]his is tantamount to saying that the dimensions of the corresponding conscious experience cannot be any less than the dimensions of the corresponding neurophysiological state. Dennett effectively removed this limitation by suggesting that even the dimensionality of the phenomenal contents need not match that of the neural vehicles. And into that epistemological crack, Dennett slipped the entire world of conscious experience like a magical disappearing act, where it is experienced but does not exist" (Lehar 2003, 379).

If content is understood as something other than the vehicle, we deviate from the materialistic world view. Lehar concludes that, "the only epistemology that is consistent with the modern materialistic world view is an identity theory" (*ibid.*). A representationalist theory that defines representation as external or by the possibility to misrepresent necessarily turns out to be a "nomological dangler" and, in Lehar's words, "corrupts the concept of representationalism".

Representationalism about phenomenal consciousness in its externalist form familiar from Dretske and Tye also strongly relies on the distinction between vehicle and content. Prinz (2004) has pointed out, however, that indication relations needs some degree of isomorphism between vehicle and object, at least on the detector side. Even if every aspect of the object need not be represented, detection has to be based on at least one specific (essential or typical) property. In the case of qualia, this argument becomes rather suspect: what isomorphism could there be between colors and detectors? This is a difficult question but my hunch is that there should be some systematicity between the brain and its ability to discriminate colors. It is likely that only brains with certain structures can differentiate colors in the same way (see the issue of the inverted spectrum discussed earlier in this section).

In my view, Lehar is too harsh towards externalism. To say, that a representation can have no relation whatsoever outside of itself, is to deny from representation its essentially intentional character of being about or directed to something outside of itself. I think that the problem Lehar points out does not

apply to representationalism as a whole. As was discussed above, some form of isomorphism between vehicle and content seems to be necessary at some level (Prinz 2004a; O'Brien & Opie 1999; cf. also criticism of multiple realizability by Bechtel & Mundale 1999).

Consider also the inseparability of syntax and semantics. If semantics and syntax are intimately related so that syntactic structures cannot be individuated without fixing some semantic relations, there seems to be extra reasons to accept a multilevel approach to ontology without embracing Cartesian dualism. Lehar's suggestion of isomorphism is also dependent on some kind of intentional relation. How could one individuate the internal isomorphic structure of a representation if not *vis-à-vis* its external counterpart? The relation between them becomes arbitrary with no pre-existing semantic relation. The same could perhaps be said of *qualia*: conscious experience is perhaps inseparable in the first place from its physical substrate. It is interesting, however, to ask whether, to resist conceivability of zombies, individuating the content of representation has to be thought in terms of identity with the vehicle. I think that this is not necessary but that it is possible to say that the individuation of content is both bound to and defined by the vehicle but also by the external fact (and for phenomenal states at least by the vehicle, O'Brien & Opie 1999). Perhaps, the vehicle could be approached without subscribing to the whole picture of representation, but then it would cease to be a vehicle of representation, and *mutatis mutandis* for phenomenality.

Externalism has a strong appeal, not least due to its close tie to causality and its idea that content must be something worldly that guarantees truth. Proponents of the narrow content insist that representational content relevant to psychological explanation does not have to worry about truth-value, in which case it threatens not to be genuine content at all. Narrow and wide contents form two aspects of meaning that are hard to reconcile. On the one hand, both accounts include something that one does not wish to abandon. One way to cope with this is to show that the incompatibility is illusory, and that perhaps we can have our cake and eat it. This is my purpose after setting the stage and making clear both views. Eliasmith concludes that no attempt at reconciliation has managed to knit narrow and wide contents together in a satisfactory manner³¹ (Fodor & Lepore 1992; Eliasmith 2000). In my view, this is better yet

³¹ These problems have been a reason to look for an alternative to representationalism. Some direct realist are so radical externalists that they think it is possible to account for perception and action in terms of relations of agents and their environment without any appeal to representations. What we are dealing with are not copies of things but the things themselves. One good example of this view is the "ecological" approach to perception of the psychologist James Gibson (Gibson 1979). In this approach, the emphasis is on the relation between an organism as a whole and its environment. Perception and other forms of interaction with the environment rely on the information that the environment already offers. Gibson takes organisms not to first recognize objects in the environment through a cognitive process and then attribute some value to them with regard to action in another cognitive process as the cognitive sequential model of that time did. Instead, he sees organisms as interacting with objects in the environment that already have some relevance for their actions. Gibson introduces the term "affordance" in this context to denote the inherent property of the environment to afford something for the organism. A path affords an

than mere internalism or externalism, and there are better prospects for such reconciliation than for a solution to the traditional problems of internalism or externalism. One way of conceiving their relation is to conceive of narrow content as partly constituting criteria for correctness, something that can be right or wrong. Narrow content is the evaluated part, and what decides whether it really is right or wrong, in turn, is the wide content.

One of the most detailed and influential arguments to view internal and external dimensions as distinct but equally real dimensions of content has been recently made by David Chalmers in a series of articles (in particular on representationalism, see 2004)³². Trying to avoid any technicalities, he sees the two dimensions of content as *epistemic* content (narrow) and *subjunctive* content (wide). A representation can have both content dimensions, and it is the problem at hand that determines which one of them is the more relevant. Subjunctive content is the ordinary externalist component of content and less problematic. Chalmers argues that we can also make a positive case for an epistemic content that is not tied to the environment but instead to an “epistemic space” that depends solely on the dimensions of ideal rational inference and cognitive significance that, in turn, depend on the inner world of a subject, and that they are the same for intrinsic physical and functional duplicates. Mental states have narrow contents that depend on this epistemic space by hypothesizing the epistemic scenario to form the actual world.

The argument bears some psychological plausibility, but it is mostly epistemic and determined by factors such as ideal rationality that outstrip the psychological capacities of the subject. Therefore, it is questionable whether it is narrow enough to serve psychological explanations. It is also assumed to apply to the contents of phenomenal experience, although it is not quite clear how phenomenal experience can relate to aspects of ideal rationality, such as hypothesizing and inferential relations. If it is possible to make such an assumption, then the idea could explain why distinct blue tits can seem indistinguishable in experience, and the behavior of the subject based on that

opportunity to advance, a stone an obstacle to advancing; a cave or the shade of a bush provides shelter and ground to lie down on, a carcass an opportunity to feed; a predator poses an immediate threat of being fed on for the organism. The organism acts in its environment according to the affordances it has as a whole, and perception is therefore only the relation between a whole organism and the environment. Naturally, this implies, and Gibson does not hesitate to draw the conclusion, that sensory processing is not necessarily relevant to perceptual psychology. According to the ecological theory of perception, perceptual processing occurs somehow in the world instead of the brain.

This kind of radical externalism has of course serious drawbacks. Should we really forget all about sensory processing, and by the same token cognitive and neurosciences? The answer is of course not. This kind of direct perception cannot provide any viable explanation of visual hallucinations such as afterimages or the possibility of introspection, not to mention such an everyday phenomenon as dreaming. In the ecological theory, perception is “nothing short of magical” (Lehar 2004, 381). Modern advocates of the ecological theory resist the most radical implications of the original Gibsonian account but have not provided any serious alternative of how to deal with its problems (*ibid.*).

³² There are many other attempts to combine the two seemingly opposite components, e.g. Prinz (2000), and one can be found even in Neisser (1976).

experience. This view manages to tie the two dimensions together in that narrow content is a condition of satisfaction for extension and serves a mapping function.

It is perhaps surprising that Chalmers uses this framework to defend the metaphysical possibility of zombies. He insists that, in the case of consciousness, epistemic content and the subjunctive content are the same. Both are determined (solely) by phenomenality that in turn is accessible from the a priori perspective (cf. Kripke 1972). A posteriori considerations would be irrelevant, since they would yield the same outcome. So, if our a priori concept of phenomenality allows us to conceive of a physically identical duplicate without consciousness, then it is a genuine possibility. (Chalmers 1996, 131–134)

The phenomenal concepts strategy seems to suggest just the opposite. It means depriving conceptual analysis of such metaphysical conclusions by arguing that the conceivability of zombies is after all compatible with materialism. However, Chalmers has argued in detail why this strategy either cannot explain our epistemic situation or give a materialist account of the possession of phenomenal concepts at the same time (Chalmers 2004), and elsewhere that a certain kind of ideal conceivability does entitle one to derive metaphysical independencies (Chalmers 2002a). He maintains that “defeating ideal conceivability will require an a priori entailment from physical to phenomenal, which will require an analysis of phenomenal concepts that can support that entailment” (*ibid.*). He thinks that thought experiments like the knowledge and zombie arguments block such analyses in principle, such that no future empirical finding can refute them, and thus one is entitled to conclude in favor of metaphysical independence. Block & Stalnaker (1999) give two reasons for doubting this. They argue, for one thing, that conceptual analyses do not suffice to ground such dependencies, since they cannot exclude “ghost” properties, additional non-physical properties that could accompany physical ones. This may be contentious, but secondly and more importantly, they maintain that no conceptual analyses are needed to establish dependences like that required by Chalmers.

In the light of the methodological considerations presented above, the primary problem with the neo-rationalism advocated by Chalmers is its defective view of concepts (see next section). Conceivability should depend on the nature of our conceptual capacities. The kind of conceivability appealed to by Chalmers is epistemic in the sense of ideal rationalism, but such views neglect the psychology of concepts and rely on the classical account of concepts as sets of necessary and sufficient conditions. To be fair, Chalmers along with Frank Jackson does not take conceptual analysis as the classical kind owing to the existence of obvious counterexamples (see next section). Instead, conceptual analysis is something like the following:

“When given sufficient information about a hypothetical scenario, subjects are frequently in a position to identify the extension of a given concept, on reflection, under the hypothesis that the scenario in question obtains. Analysis of a concept proceeds at least in part through consideration of a concept's extension within

hypothetical scenarios, and noting *regularities* that emerge. This sort of analysis can reveal that certain features of the world are highly *relevant to determining* the extension of a concept, and that other features are irrelevant." (Chalmers & Jackson 2001, 322, emphases PL)

"The possibility of this sort of analysis is grounded in the following general feature of our concepts. If a subject possesses a concept and has unimpaired rational processes, then *sufficient empirical information* about the actual world puts a subject in a position to identify the concept's extension." (Chalmers & Jackson 2001, 323, emphases PL)

Not much, in fact, hangs on whether this view of conceivability is adequate or not. The negative aspects of traditional conceptual analysis extend, in my view, to the account given by Chalmers and Jackson. They pre-exclude the effect of conceptual evolution in their account. In this citation this is secured by the term "sufficient information". This is controversial in its own right, but more important is the question whether we really do possess the sufficient information about phenomenal consciousness. Chalmers and Jackson continue to argue like this based on their own intuitions. Chalmers thinks that we are in such a position that no amount of extra information about the physical substrate of phenomenal consciousness could impose such constraints that would change our dualist intuitions. In my view they have not managed to show that the intuitions on which their view is based cannot be false or change. No "sufficient empirical information" is backing up their argument. Noting regularities in our current intuitions with respect to fantastic scenarios is not the same as having sufficient empirical information about the world. At present, it may seem to many that there may be variation in phenomenal consciousness regardless of its physical substrate. However, many think that it is a perfectly possible scenario that our concept of phenomenal consciousness as not able to vary regardless of its physical substrate can become fathomable given more and more information about the physical substrate.

All this ultimately comes to whether we should accept the strategy proposed by Chalmers and Jackson, which is based on conceptual entailments, as our guide to explanatory issues. I believe that criteria of good explanation are not whether we can deduce *a priori* phenomenal facts from physical ones as required by Chalmers and Jackson. Nor does it even matter whether such hypothetical scenarios can be formulated. What is more relevant is that we can "see" how phenomenal facts work by virtue of their physical substrate. Once this is achieved, our concepts can eventually be changed and advance toward the interdependence of phenomenal and physical states. We would no more believe in the possibilities of such scenarios (cf. Van Gulick 2003). Regularities in *a priori*-formulated hypothetical scenarios are not immune to conceptual evolution. The source of conceptual evolution is mainly in the findings of the empirical sciences that narrow the epistemic space, or, in co-evolutionist terms, constraints are imposed on the meanings of the concepts at other levels. This non-stagnational view of concepts is also more compatible with the recent

psychological research on concepts, and it also seems to be one favored by Ned Block and Robert Stalnaker:

“We argue that hypotheses about ghost properties are ruled out on empirical methodological grounds, rather than by conceptual analysis. The same kind of methodological considerations might be used to argue against dualism about consciousness.” (Block & Stalnaker 1999, 19–20)

“Closing of the explanatory gap in the case of life has nothing to do with any analytic definition of “life,” but rather is a matter of showing how living things around here work.” (Block & Stalnaker 1999, 15)

I concur with Block and Stalnaker. Their account is in line with mechanistic explanations and the co-evolution view endorsed above, although these concepts do not figure in the paper by Block & Stalnaker.

3.3 Representational content: conceptual or non-conceptual?

The long tradition of the view in analytic philosophy holding only conceptual content as intentional or representing something has been losing ground. Representationalism about phenomenal consciousness is an example of this. Both views, however, usually accept the division of the representational mind into two basic kinds of representations, conceptual and non-conceptual. There are of course many ways of making this distinction, in addition to the traditional one based on linguistic capacities: fineness of grain (e.g., Evans 1982), and spatial contents (Peacocke 1992). Of representationalists, Dretske analyzes this distinction by reference to an analog–digital distinction: analog representations are continuous while digital ones are discrete (Dretske 1995). They all share the idea that the non-conceptual contents of perception *outstrip* our conceptual capacities: we do not have concepts to categorize all the aspects of perceptual content. In addition to contents, a corresponding division can be made at the level of representational capacities (or states, see Heck 2000).

In order to determine whether content is conceptual or non-conceptual, one should have a pre-conception of what ‘conceptual’ means (Byrne 2005; Bermúdez & Cahen 2008). There are many conceptions available both in philosophy and psychology³³, and we can turn to different theories in order to see what we ought to say about the conceptual and non-conceptual realms³⁴.

³³ Good introductions to both philosophical and psychological studies of concepts are given in Margolis and Laurence (1999) and Prinz (2002). The first contains classical psychological studies, but for a more comprehensive review of psychological studies and theories, see Murphy (2004).

³⁴ Byrne (2005) remarks that the conceptual–non-conceptual debate never concerns psychological views of concepts. My interest is primarily in psychology and because higher-order theories rely on psychological capacities, I will concentrate on the relation between phenomenal consciousness and concept possession. Moreover, I think that concepts constitute yet another good example of why conceptual analysis and empirical data must constrain each other.

This move is important for two more reasons. First, in order to distinguish between different kinds of first-order and higher-order representational theories, any knowledge about the nature of the conceptual and non-conceptual is invaluable. Second, there are methodological consequences. Conceptual analysis is a major method of philosophy, and also of the present study in a certain sense (see chapter 1). How conceptual analysis can properly proceed or what we should make of the outcome depends largely on what we take concepts to be.

Traditional theories of concepts come in two flavors (Prinz 2002). Since Plato, it has generally been taken for granted that concepts are *definitions*, usually understood as sets of necessary and sufficient conditions. This view, often referred to as the “classical view”, met with opposition along with the rise of empiricism in the modern era. Empiricists famously held a “image theory” of concepts (or “ideas”). Ideas stem from initial impressions which themselves are mental images, copies of singular sensations. Ideas are mental images as well, but they differ from impressions in that they purport to apply to many instances of referents. An idea is a picture of general features³⁵. There are advantages to this view, especially in its account of how we acquire our concepts. Image theory has enjoyed a recent resurgence in some of the contemporary theories of concepts (*ibid*; Barsalou 1987, 1993)

The classical theory of concepts was dominant for centuries. The origin of the contemporary theories of concepts lies largely in the shortcomings of the classical theories. These shortcomings were mostly revealed by the experiments conducted by Eleanor Rosch (1975)³⁶. Rosch showed “overnight” that the classical theories were not only conceptually flawed but also psychologically implausible. The psychological ability of categorization is not manifested as the application of necessary and sufficient conditions. Instead, it is manifested as “typicality effects”. This means that one categorizes typical things in a faster and more reliable way. A robin is recognized as a bird faster and more reliably than an emperor penguin. This is not predicted by the classical view, since it does not treat some necessary and sufficient conditions as more salient than others. Furthermore, features that are neither necessary nor sufficient may influence categorization. Therefore, concepts must be something else than definitions. The typicality effect is something that a theory of concepts worth its salt must handle. Rosch suggested that concepts are more like *prototypes*, sets not of necessary and sufficient features but of typical features that are do not have to conform any set that is common to all members but may form a set that bear a “family resemblance” relation to one another.

³⁵ This may sound to many like a *contradictio in adiecto*. Early empiricists saw this, and attempted to explain this by the method of elimination: One starts from an image of one instance in a category, e.g. a picture of an individual robin. This image is then compared to other instances and the features that varied across the compared instances are eliminated, leaving only those features that were common to all instances. An image of these features can thus represent the features of several instances.

³⁶ In philosophical studies of concepts, Putnam and Quine are the most influential critics of the classical view (see articles in Margolis and Laurence [1999]).

The prototype theory, nevertheless, has a hard time explaining intentionality, which is another important property of concepts. On the prototype theory, it is not less than a mystery how a set of changing and loosely connected features can account for the property of concepts of referring to a determinate set of things. Nothing that contradicts intentionality is entailed by the prototype theory, but it simply has little to say about the intentionality of concepts.

The “exemplar view” is another theory based on typicality. It is basically an image theory version of the prototype view and is sometimes considered as a specific version of it. Here, the question of perceptual and non-perceptual representations is important, and therefore these two views are best kept apart from each other. An exemplar is a mental image of a specific instance of a category. On the exemplar view, concepts are images or compilations of images of specific exemplars. There are also psychological reasons to keep the two views distinct. There is experimental data specifically supporting the exemplar view over the prototype view and *vice versa* (Murphy 2004). This may support best the view that both exemplars and prototypes exist as real psychological phenomena. The exemplar theory suffers from the problems of both the prototype and the picture theories. It cannot serve as a theory of concepts as it stands. Nevertheless, I agree with Prinz (2002) that exemplars exist as mental structures and they affect categorization, while on the other hand they cannot be identified with concepts.

The third view of concepts is once more a non-imagistic view. The theory theory view is already familiar from the discussion of functionalism and will be discussed later in connection with our ability to understand the minds of others and ourselves. Its basic tenet is that concepts can be spelled out with the use of the concept of theory. Whether theory theorists think that concepts are mini-theories themselves or parts of theories determined by theories is not spelled out (Prinz 2002; Murphy 2004). Theory theory is a holistic account: concepts are at least partly constituted by their relations to other concepts. Our background knowledge affects categorization (the so called “knowledge-effect”, also reviewed extensively in Murphy 2004); hence it has been suggested that categorization and concepts consist of networks of knowledge. For example, a layperson naturally categorizes a robin-looking bird as robin, but when given the information that someone has painted a blue tit to look the same as a robin, people change their categorization accordingly.

Akin to internalist views of content, the theory theory cannot easily handle intentionality. If our knowledge structures are not stable, how it is possible to refer to stable things in the outside world (as philosophers often suppose there is)? The theory theory is also held to lack an account of compositionality because of its leaning on holism (Fodor 1998). Fodor’s own theory, *informational atomism*, is an externalist theory that relies on things in the outside world and has been developed to handle intentionality and composition. Fodor’s theory is, however, implausibly atomistic and commits itself to a strong innatism, since it cannot allow other concepts to affect the acquisition of basic concepts.

Overall, specific theories of concepts are suited to handle some of the alleged features of concepts while they have trouble with others. Consequently, there might be reason to suppose that if no single theory accounts for all the aspects of concepts, a combination perhaps will. Specific combinations have been presented, but as Prinz points out, perhaps a more generic and unified view would be better than an aggregate view. Prinz's own *proxytype theory* pretends to be such a view. The proxytype theory is a partly imagistic, perceptual theory suggesting that specific exemplars are stored from confrontations with them. These are the conscious part of concepts. There is, however, a non-conscious part consisting of large non-conscious networks of background knowledge. Together, these components account for the typicality effects (exemplars) and knowledge effects (networks of non-conscious background knowledge). Both have trouble with intentionality. Prinz secures this by maintaining that concepts are *detectors*, internal structures that allow an indication relation. Indication does not require isomorphism: a simple structureless flashlight can indicate the presence of a cow. There has to be, however, something that makes this possible and this is explained by detectors, mediating components that can track things in the outside world with respect to one or many properties. These components possess a degree of isomorphism to the property they track. Note that isomorphism does not necessarily mean perceptual resemblance but it can be in a coded form and thus not directly visible. Through the detectors, the proxytypes are in an informational relation to things. Therefore, the proxytype theory can enjoy the full privilege of the theory of intentionality of informational atomism. One might wonder how it is possible that the mediating components, notably those of externalist informational relations and internalist holist aspects can co-exist. The answer lies in the two-component theory of content held by Prinz (Prinz 2000a) that is akin to two-component theories of content that were favored above.

I think that the direction in which Prinz takes his theory of concepts is the right one. However, in addition to some complications with his theory of content, there are certain phenomenological aspects that are suspect in my view. Namely, in the light of this account, the whole non-conceptual-conceptual distinction becomes questionable. Prinz argues that there is no difference in principle between perception and conception. All contents are governed by criteria of satisfaction in the same way and admit only of degrees in the dimensions of abstractness, fineness of grain and generality. I think that this argument may succeed in downplaying the differences between conceptual and non-conceptual content and abilities, but not nullify them altogether. The line can be drawn, in my view, phenomenologically, relying on differences in such conscious phenomena as mental imagery and concepts. Since images cannot capture a general content and a general content does not phenomenologically look like anything, there must be a difference between the two; in experience we do confront both types of content. Note that Prinz holds that general knowledge must be non-conscious. This view is misleading in that it occludes the non-imagistic phenomenology of general concepts. There is a difference at the level of experience between a mental image and a concept,

though concepts are often accompanied by related mental imagery. If there were no difference, a sole mental image and a concept accompanied with the same mental image would appear the same.

Consider the example of Bermúdez (2008): a 45° tilted square can be perceived as a tilted square or as an upright diamond. What makes the phenomenal difference in seeing the object as either one is something that must outstrip the spatial and other imagistic (non-conceptual) content of the object, since they are the same for both of them. There are a number of organizational elements of perception that spontaneously mold the *phenomenal* properties of the object even if the object remains the same in imagistic terms. One way to characterize them is to view them as “higher-order principles” of perception (Bermúdez 2003). These include grasping an object as a whole body, grasping the cohesion of the object, the tendency to follow a certain trajectory when moving etc. Developmental psychology has shown all these to be present already in non-linguistic infants (Spelke 1990). These elements are also recognitional abilities, and we do feel the phenomenological difference they make in perception³⁷, even the sensation of familiarity if we confront the same object again. Bermúdez lumps all this in with the notion of non-conceptual content, since he takes concept possession to be something that requires the ability to reflect on the grounds of the concept, which he takes to be an essentially linguistic operation (e.g. Bermúdez 1998, 68-71). His criterion differs from mine and many others. It is important to note, however, that the higher-order principles espoused by Bermúdez do not differ much from recognitional capacities, which is for many the mark of the conceptual abilities. Most often when concepts are thought of as non-linguistic they are thought of as purely recognitional capacities (Loar 1997; Carruthers 2000a; McDowell 1994a). This provides the criterion for a concept to be a concept that it be an ability to discern some thing at time t_1 and, when confronted with it at a later time t_2 , to identify it as the same thing. No language is required in this operation because that thing could well be something that we have no word for, and this ability surely is present in aphasics and animals.

One could object that higher-order principles might contribute to the non-conceptual representational content of perception only for the time the perception persists and then disappear. But this is not how Bermúdez defines higher-order principles. Higher-order principles contribute to the perceiving of objects as whole particulars. He requires them to be context-free and applicable in conditional thinking. As such, they cannot persist only during the time of perception. Now, this would not make sense if all these operations were built anew every time we have an experience of the same object.

In short, the conceptual sphere could be thought of as differing from the non-conceptual sphere in that non-conceptual content makes up fine-grained perception-like representational content and conceptual content abstracts some fine-grained non-perception-like higher-order features from the non-conceptual sphere. As such conceptual capacities constitute capacities of individuating,

³⁷ Recall the debate between Titchener & Külpe (n. 10).

categorizing and recognizing whole objects as particulars. Concepts conceived in this manner also provide a vehicle of thought that might both exist at personal and sub-personal level, and can thus provide the vehicle of thought that allows metacognition. The problem of metacognition in non-linguistic creatures would not, then, be whether they possess suitable vehicles that allow metacognition but rather whether they possess adequate psychological concepts.

The nature of conceptual-nonconceptual distinction is, it might be objected, a thing that one cannot settle with phenomenology since it is the source of the greatest dispute in the history of introspectionist psychology. I disagree: surely phenomenology can make advances and a dispute should not call a halt to the entire enterprise. Concepts and mental imagery are experiential phenomena, and I believe that the interplay between phenomenological, conceptual and empirical studies has advanced and led to an accumulation of knowledge on this issue, and that this will continue in the future.

3.4 Evaluative intentionalism

Murat Aydede (2001) has raised an important counter-argument against the ability of the representationalist theory of Dretske and Tye to handle an important class of conscious mental states, namely *affective* states. Many affective states are experiential states that seem inexorably mental, such as pain and emotions, whose contents supposedly do not depend on external objects. *Pain* is often taken as the most compelling instance of this. In the philosophy of mind, it is often even taken to be a paradigm example of a state that is only determined by its mental or phenomenal properties (Kripke 1972). It does not have to be directed at any physical properties of the world. *Prima facie*, pain and emotions seem to form the greatest obstacle to intentionalism. It is hard to see what the intentional object of an emotion is supposed to be. A pain is a pain even without a tractable physical origin. The transparency of mind does not seem to extend to affective states. Emotions may have objects; the object of my love is the person I am in love with. In contrast to sense experiences, however, love seems to involve experiential qualities that do not involve the object of my love. The experiential feeling is difficult to identify as a property of one's loved one.

Intentionalists have not been discouraged by this counter-argument (Tye 1995, 2000). The main strategy in answering it has been to draw on the Jamesian somatic theory of emotions recently defended by Antonio Damasio (1999). The somatic theory of emotions assimilates emotions to bodily perceptions. Emotions represent configurations of bodily changes specific to each emotion. Representing involves criteria of correctness: one can be right or wrong about one's emotions. Pain without a tractable bodily origin is pain that misrepresents; love without bodily perception is imaginary love.

There is, however, a deficiency in this kind of perceptual analysis of emotions. A somatic perception does not provide any criteria of correctness about the evaluative, naggingly negative element essential to pain or the deeply warm and cherishing element toward its object characteristic of the feeling of love. Murat Aydede (2001) has elaborated this idea into a full-fledged critique of representational theories of emotion. His analysis primarily concerns pain but he says it can be generalized to concern emotions as well (p. 67). Aydede contests the intentionalist account, using pathological data from pain-research. Certain patients suffering from severe chronic trigeminal neuralgia (unbearable continuous pain in the facial area) were treated by a certain kind of lobotomy where the connections between certain somatosensory areas and limbic system were dissected. These areas are presumably responsible for the *sensory* and *evaluative* components of pain. Dissection of these areas resulted in a massive improvement in the condition of the patients. They reported still feeling the pain – it just did not bother them anymore. Translated into the terms used previously, the sensory component was intact but it no longer activated the evaluative component.

The moral drawn by Aydede is that intentionalists can only deal with the sensory component of pain, which suffices to represent the state of the body but they do not have any reasonable account for the evaluative component of pain. By parity of reasoning, as emotions are inherently evaluative – they have a valence of being positive or negative – mere representation of the body is insufficient at best. Something more is needed to explain the evaluative component of emotions. As a solution to this acute problem, Aydede sees a reversion to some kind of psychofunctionalism, to impure representationalism (see n. 26) as the only alternative.

In the case of affective states, intentionalism seems to violate the first clause of Dretske's representational thesis (p. 44), that all mental states are representational. Moreover, it has a hard time with the second. The somatic theory of emotions does not offer any plausible informational function for emotions: "Emotions promote behavioural responses. We run when we are afraid. If emotions represented bodily changes this would be unintelligible. Why should we flee when our hearts race?" (Prinz 2004a, 59). The relevant informational link to the environment, to the objects of emotions seems to be lacking.

Representationalism, then, needs something else to deal with emotions. Recently, Jesse Prinz (2004; see also Seager 2004) has developed a promising way to deal with the evaluative component in terms of representationalism. I call this strategy *evaluative intentionalism*.

Instead of representing bodily changes only, Prinz argues that emotion represents the *relation of the organism to its environment*. Emotion represents the relation of the organism to objects and states of affairs external to it that are significant for its well-being. Like pain, emotions are essentially evaluative. Borrowing the terms coined by cognitive theories of emotions, Prinz sees emotions as "appraisals" that represent "core relational themes". Core relational themes are emotion-specific evaluations of the relation between the

organism and its environment. The core relational theme of fear, for example, is a possible threat to the organism.

Prinz's theory does not, however, abandon the somatic component. Confronting something dreadful causes somatic changes that prepare one to act accordingly. Somatic changes are registered by the organism. By registering the somatic changes, Prinz proposes, the organism can obtain information about what sort of situation it is confronting. Emotion has an informative function about the situation the organism is in: a small message carries a larger representational content about the environment. By registering the somatic configuration elicited by the situation, the organism represents its relation to the environment. Thus, Prinz calls emotions "*embodied appraisals*" or "*valent perceptions*" (p. 229). An evaluative component is intrinsic to emotion but in representational terms. The appeal to psychofunctionalism becomes unnecessary. At the neural level, the evaluative component is also representational. Pain is representation of a bodily state as harmful. If it does not feel bothersome anymore, the failure is in one's representational capacities, not in the psychofunctional role of the sensory representation.

In addition to Prinz and Seager, Timothy Schroeder (2004) has independently formulated a theory of desire that shares the general form of evaluative intentionalism. Desire is an important element of analytic philosophy, especially action theory and the philosophy of mind. Schroeder points out that, nevertheless, proper theories of desire continue to be lacking, despite complaints about how acutely one is needed (p. 3–4). Two traditional views, nevertheless, stand out. Schroeder calls these the "*motivation theory*" and the "*hedonic theory*" of desire. The first can be found in the tradition of action theory, and the latter can be found in the work of Galen Strawson (1994). Schroeder's strategy is such that he first gives the conceptual and then the empirical reasons for which to rebut both theories. Then he goes on to outline an alternative theory: the reward theory of desire.

The motivational theory holds simply that desire is anything that motivates to action. In action theory, mind is viewed as containing representational elements, the propositional (conceptual) contents that can be given in the form of a that-clause, but also non-representational functional elements, attitudes towards these contents. The attitude one has toward the propositional content depends on what kind of role the propositional representation occupies in the functional system of the mind. If it occupies a role in which it specifies what one wants to bring about in the world, then the attitude is motivational (desire). Its "*direction of fit*" is that the world should accommodate to it. If, in contrast, the same representation with the same propositional content occupies a role in which it steers the action, then the attitude is veridicality (belief). Its direction of fit is that it should accommodate to the world. On this view, anything that occupies the motivational role is desire by definition.

On the hedonic theory of desire, motivation plays no essential role. Instead, desire is something that causes or is disposed to cause pleasure. Galen Strawson (1994) has tried to show that mind does not necessarily depend on

action at all (recall “Weather Watchers” from section 3.1.). Nor should thus desire depend on action. Beings that do not act at all can have dispositions to experience pleasure, and be claimed to have desires in this sense. Changes in weather can cause pleasure and displeasure in these creatures. On the hedonic theory of desire, they can have desires in the sense of mental states that anticipate such changes in weather, which give rise to experiences of pleasure.

Schroeder accepts the conceptual critique of the hedonic theory on the motivation theory of desire. He adds that the dissociation of action and desire and pleasure is, in fact, a double dissociation, established by conceptual and empirical research. There are motivational states that we would not want to call desires, such as habits. Moreover, he claims that there are actual empirical cases in which action is prohibited but mental states, pleasure and desires are not, and vice versa. An example of this is Parkinson’s disease. In Parkinson’s disease, volitional action becomes absent, apparently because of a lack of inhibition normally executed by the basal ganglia. The inhibition does not occur, since, in Parkinson’s disease, the striatal connections that feed this system are absent due to the decrease in the number of the dopaminergic neurons of the striatum.

The hedonic theory itself does not fare any better whether measured conceptually or empirically. In some pathological conditions pleasure is absent, but desire is not. This becomes clear once desire is defined correctly. Schroeder maintains that instead of connecting it essentially with either pleasure or motivation, we should think of desire in terms of *reward*. By reward he specifically means something that is manifested in one kind of learning, namely, instrumental conditioning. In instrumental conditioning, certain behavioral patterns are reinforced when they are rewarded. The most classical example of this is the pressing of a lever by a rat that is then rewarded for so doing, e.g. by receiving food. The reward induces the rat to press the lever all the more frequently. Schroeder maintains, however, that reward is not necessarily related to pleasure. This is because the brain structures related to reinforcement learning (the ventral tegmental area/substantia nigra pars compacta of the basal ganglia in concert with the orbitofrontal cortex) are distinct from those related to pleasure (the perigenual area of the anterior cingulate cortex). Reinforcement learning is not affected by lesions in the orbitofrontal cortex, and rats whose brains have been connected by scientists with the areas associated with reinforcement and which can directly stimulate these areas by lever-pressing, end up doing nothing but pressing the lever, even though the signal totally sidesteps the pleasure areas. Avoiding a punishment also elicits the reward signal, and at a conceptual level one can say that one may want to avoid a punishment although this need not involve any pleasure.

What is crucial about the brain areas associated with reward is that they are the mechanism for the releasing of dopamine elsewhere in the brain (there are dopaminergic connections to nearly every part of the brain). The presence of dopamine contributes to the strengthening of neural connections, in other words, to learning. At a phenomenal level, one’s attention is directed towards something that sends reward signals more often, both in terms of doing it more

often and of thinking about it more often. Schroeder suggests that this is best viewed from the standpoint of representationalism: desiring something is representing it *as a reward*, defined as something contributing to reinforcement learning. In this way, desire does not form an obstacle to representationalism as an explanation of consciousness³⁸.

Evaluative intentionalism opens up an interesting perspective on affective and evaluative states in representational terms, including conative states such as desire³⁹. This perspective is also of interest also with respecty to higher-order representations, a topic that neither Prinz or Schroeder deals with systematically. First, however, the higher-order approach needs to be introduced.

³⁸ Schroeder, however, holds that reinforcement learning occurs non-consciously, and that we are only able to become conscious of side-effects of reinforcement learning, e.g. the urge to act or subsequent pleasure. In my view, this is controversial: there may well be some phenomenology related to reinforcement learning.

³⁹ Prinz (2002) argues that it is clear that action also can be explained in representational terms. I shall examine this suggestion in chapter 5.

4 HIGHER-ORDER THEORIES OF CONSCIOUSNESS

4.1 First-order and higher-order representationalism: a first approximation

Representationalism as just outlined admits of an important further distinction. Each sort of representationalism includes the possibility of metarepresentation or higher-order representation.

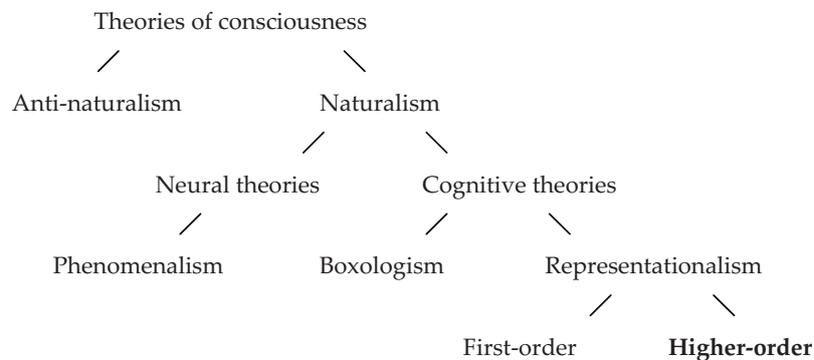


FIGURE 4 The consciousness tree, version 4

A *higher-order* representational theory of consciousness holds that metarepresentational capacities are essential for consciousness⁴⁰. Consciousness, then, is based on self-awareness, or more precisely, awareness of one's mental states. In this chapter, I will present and discuss these theories. First, I will consider the advantages that bringing in self-awareness and a

⁴⁰ It may be possible to concur with phenomenism that phenomenality does not have to do with semantic or representational properties at all. However, as far as higher-order awareness of one's mental states is concerned, avoiding the issue of semantics may be unavoidable.

metarepresentational dimension have for representational theories of consciousness.

The first and foremost reason for endorsing a higher-order account is to discern conscious states from non-conscious ones, as it was already mentioned in the introduction in the form of the transitivity principle (TP):

(TP): "A mental state is conscious only if one *is in some way aware* of it." (Rosenthal 2006, emphasis PL)⁴¹

The main reason for accepting the transitivity principle is the alleged observation that this is quite simply what we mean by consciousness. If a layperson were asked to describe what makes the difference between conscious and unconscious mental states, something along the lines of the transitivity principle would be the answer he would give. Amie Thomasson calls this kind of evidence "verbal evidence", but points out that other related kinds of evidence have been presented, namely "phenomenological" and "epistemological" evidence (Thomasson 2005). Phenomenological evidence is the marginal phenomenal givenness or self-awareness of consciousness in the experience (Kriegel 2003). Epistemological evidence is that consciousness is "first-person knowable", available directly only to the person to whom that consciousness belongs. Thomasson, however, points out that verbal and phenomenological evidence are subordinate to epistemological evidence. Our way of speaking about and observing our experience depends on our epistemic access to it. The dependence of every piece of evidence on epistemological evidence in this way, in fact, makes all of them controversial. Surely, this kind of consciousness is the type we become aware of in the first place *per definitio*. Yet, it is controversial whether how we know about consciousness is the way consciousness really is. There may well be consciousness that we are not or cannot become aware of. No reason other than some conceptual intuitions is given for *why* the way for a mental state to become conscious is to be aware of it.

If the main motivation for endorsing a higher-order theory is to be able to distinguish between conscious and non-conscious mental states, a natural way to render such a theory controversial would be to provide an alternative, better criterion for discriminating between conscious and non-conscious mental states. In this case, we would have a reason to abandon the higher-order representational theory in favor of a first-order representational theory or some other theory of consciousness. In my view, there are ways of explaining the difference between conscious and non-conscious mental states in first-order terms, as in terms of global availability (see section 2.4).

⁴¹ One could add the flipside of this principle that mental states we are not conscious of are those that remain non-conscious. Higher-order theorists maintain that this is a genuine problem for the first-order views, although it is questionable why first-order theories should not succeed in that by appeal to global availability (recall "poisedness" in Tye 1995).

The second motivation for endorsing a HO view is the alleged incapacity of the first-order theory to distinguish world from its appearance (Carruthers 2000). Making the *is-seems* (or appearance–reality, further often “A–R”) distinction requires cognitive capacities that cannot just concern the world; the cognitive capacities required must be of a higher-order. It is true that the is–seems distinction figures in the experience of a normal adult human subject. However, it can be questioned whether the is–seems distinction is essential to conscious experience. HO theorists insist that it is, since it follows directly from the transitivity principle. Denying the principle would relieve one of that burden. For Carruthers himself, experience acquires its phenomenal feel and relevant subjectivity and perspectivalness from the making of the A–R distinction. If the distinction is not made, the experience remains non-subjective. This sounds highly controversial to me: how can one be aware of phenomenal properties if the experience does not already have them? Moreover, it is questionable how a mere awareness or subsequent inferential role can bestow such properties on experience. It seems to require some kind of reverse causality. It is hard to imagine examples of knowledge having similar effects on anything. Knowing hardly affects its objects in a manner that is this radical.

FO theorists Tye and Dretske claim that the A–R distinction can be made using phenomenal concepts. However, in their case, the object is not the mental state or its properties but the representational properties of the worldly objects in question. This also constitutes their theory of introspection: introspection is awareness enabled by phenomenal concepts of one’s mental states by way of being aware of objects. Introspection is “displaced perception” (Dretske 1995, Tye 2000). This is seen as problematic in two respects. It threatens the first-person authority and privileged access by making introspection resemble inference. Representationalists have replied to these worries (Dretske 1995; Aydede 2003). Personally, this flies in the face of my phenomenological intuitions; I think that in addition to the object, the manner of representation can be given and become the object of actual higher-order representations (elaborated in section 5.1). Moreover, a sort of dilemma seems to follow from this view: the applying of phenomenal concepts to objects seems to collapse into a sensory concept, since their contents are identical by definition. In this case, there would be no difference between appearance and reality. This is of course in line with the spirit of first-order representationalism and the transparency of experience (TE, section 3.1). If appearance and reality had distinct contents, first-order theories of consciousness would qualify as higher-order theories as far as awareness of mental states is concerned. Surprisingly, this intermediate alternative view has been neglected and therefore my purpose is to develop such a view in chapter 5. Since I reject the transitivity principle, I have different reasons for preferring first-order views of consciousness. For developmental reasons, I think that first-order theories suffice as far as the phenomenality of infants and animals is concerned, as infants and animals are not originally aware of their mental states in any meaningful way. This is discussed in terms of “psychic equivalence” in chapter 5 (see also Fonagy et al. 2002). As Thomas

Metzinger has urged, transparency of experience does not, in fact, reveal the representational nature of our mental states but instead hides it (Metzinger 2004, section 3.2.7). However, I think that awareness of mental states is possible after one develops the cognitive capacity to differentiate between appearance and reality (see *ibid.*, p. 163). This practical capacity is developed in the primary affective relationship with one's caregivers.

It is thus important to see that theories based on the transitivity principle have to make the following important commitment. Since these theories hold consciousness to be conscious of itself, they have to assume that consciousness must be able to *differentiate* between the objects it is conscious of and itself, the consciousness of those objects. But what is it exactly that guarantees any awareness of *experiences*⁴²? Consciousness is, of course, directed at objects but it must also be aware of itself at least as the *appearance* of its objects, of its own intentional character to constitute awareness of experiences. Otherwise, not experiences but only worldly objects would be given. I will refer to the capacity to be aware of one's experience in this manner as the appearance–reality (A–R) distinction. In order to be conscious not only of objects but also of one's being conscious of objects requires in some way this A–R distinction, the ability to distinguish between mind and the world. On the transitivity principle, the A–R distinction in this sense belongs to consciousness *as an essential property*.

Most theories holding the transitivity principle take awareness of one's mental states to be immediate and non-inferential. There is a *privileged access* to one's occurrent mental states. It does not make one conscious if one is told that one has a certain mental state or if the awareness of the mental state is acquired by inference from one's actions. Therefore, e.g., Rosenthal claims that one must be aware of one's mental state in the direct way of being aware of one's *present* mental states in a suitable way (Rosenthal 2005). Though the transitivity principle is a necessary condition for consciousness, it is not sufficient, since there are ways of being aware of one's mental states without that involving consciousness of those states. Some extra conditions and specifications about the way the awareness is realized are needed in order to achieve a proper theory of consciousness. Theories that implement the transitivity principle differ from each other in how one's awareness of one's mental states is implemented – what kind of access one has to one's mental states (Thomasson 2000, Rosenthal 2005). The classical accounts of the ways of being conscious of one's mental states are higher-order thought and higher-order perception theories, and they will be presented in more detail in the next section^{43,44}.

⁴² One could take this to mean awareness of experiences *as experiences*. And this is right in the sense that when one is aware of experiences one is aware of what they are, i.e. experiences. This does not mean, however, that one is necessarily aware of experiences as experiences in a conceptual or thematic way. I think that all alleged ways of being aware of experiences, including non-conceptual and non-thematic, cannot help falling prey to this requirement.

⁴³ There are, however, some higher-order views that do not share the view of immediate and privileged access. These theories equate the way other people and oneself know of one's mental states. On this view, one's access to one's own mind is the same as access to the minds of others, namely a conceptual theory (Gopnik 1993)

Nonetheless, all higher-order views imply that one has to have the right sort of *psychological* capabilities to be aware of mental states. Elucidation of the nature of these capacities and of one's privileged access to one's conscious states is crucial to any theory based on self-awareness of consciousness. I will try to state my opinion at the end of this chapter and continue on the same topic in the following one.

As it was described above, phenomenologists take phenomenal consciousness to be the most basic form of consciousness and access, reflective or reflexive consciousness to be based on phenomenal consciousness. In the HO theories, in contrast, phenomenal consciousness is more complicated, and access and reflexive consciousness are more fundamental. Phenomenal consciousness can be explained by a certain kind of reflexive or access consciousness that constitutes the easy problem of consciousness. However, the truth about HO theories is not so straightforward. In the last section, I will show that HO theorists are trying to explain something else than phenomenal consciousness. This something else is not very well described by them, and my own suggestion for the kind of consciousness that can be explained by HO theories is mentalized consciousness, which I will describe in more detail in the following chapter.

4.2 Two basic views: HOP and HOT

The most common version of the higher-order theories of consciousness is the higher-order perception theory (HOP). It is the classical Lockean version of the internal or inner sense view, of which a form can also be found in Kant (1778). The most widely known contemporary proponents of the inner sense theory are William Lycan (1987; 1996; 2003) and David Armstrong (1969; 1984 & 1997). The inner sense theory holds that what distinguishes conscious and non-conscious mental states is an inner perception of the state.

The traditional Lockean inner sense theory holds that there are two kinds of perceptions: inner and outer. Outer perception consists of the traditional sense modalities of vision, touch, hearing etc. Inner perception does not refer to the sense modalities that are directed to the body instead of extra-corporeal reality. Inner perception is a form of perception that is directed to the outer senses themselves: we perceive our outer perceptions.

“Secondly, the other fountain from which experience furnisheth the understanding with ideas is, -- the perception of the operations of our own mind within us, as it is

or a simulation (Gordon 1995). These theories are perhaps not usually put forward as theories of consciousness but rather as development of self-awareness and theories of mind. In any case, they deny any form of the givenness of consciousness or privileged access as well as the obvious ability of people to introspect their own mental states, and are thus out of the scope of the present discussion.

⁴⁴ Many of the issues of this debate were thematized, in fact, already in the Middle Ages (Yrjönsuuri 2007).

employed about the ideas it has got [...] And such are *perception, thinking, doubting, believing, reasoning, knowing, willing*, and all the different actings of our own minds; -- which we being conscious of, and of observing in ourselves, do from these receive into our understandings as distinct ideas as we do from bodies affecting our senses. This source of ideas every man has wholly in himself; and though it be not sense, as having nothing to do with external objects, yet it is very like it, and might properly enough be called *internal sense*. But as I call the other Sensation, so I call this REFLECTION, the ideas it affords being such only as the mind gets by reflecting on its own operations within itself. By reflection then, in the following part of this discourse, I would be understood to mean, that notice which the mind takes of its own operations, and the manner of them, by reason whereof there come to be ideas of these operations in the understanding." (Locke 1975, 125)

The Lockean theory of mind is partly a reflexive model, and Locke's own term for inner sense is "reflection". The precursors of this view can be traced to the Middle Ages and further back on to Aristotle (Aristotle 1986, Caston 2002). In Aristotle, however, inner senses (sic!) served quite different purposes, such as unification of the outer senses, imagination etc. The Lockean view has better retained its basic form up to the contemporary philosophy of mind, and it has influenced all the subsequent classical philosophers. Of the successors to Locke, Kant (1778) held that there are two forms of self-awareness: internal (inner) sense and apperception. Internal sense is the way mind reveals itself to itself. Outer perception is organised spatially whereas internal perception is organised temporally. Time is the form of inner sense: when one observes one's mind, its contents appear in a temporal succession. This view deviates in some respect from the Lockean view, but the basic idea can be viewed as essentially the same. Inner perception is the capacity of the mind to turn on to itself and observe its own contents. Apperception, the second form of self-awareness for Kant; it is the modal capacity of the mind of being able to conceive its contents as its own. According to Kant, this capacity must be inherent in all contents of the mind so that one must be able to think of any episode of thought as one's own thought⁴⁵.

Following Locke and Kant, versions of inner sense or inner perception have been adopted in many traditions. It has become a feature of everyday vocabulary to talk about an "inner sense", an "inner perception", the "inner eye", or the "mind's eye". Unfortunately, these concepts remain vague and pre-theoretic, and a full appreciation of these terms is not possible here. Elaborate theories of inner perception are far less frequent, and theories of what this capacity amounts to can be found, perhaps in addition to the forefathers and fathers of the phenomenological tradition, Brentano and Husserl, in the contemporary philosophy of mind in the debate over the nature of the higher-order representational theories, which is the main target here.

David Armstrong's recent version of the inner perception view (1968; 1984 & 1997) is the modern version of it in the analytical philosophy of mind, and it has, in turn, inspired subsequent versions of it. Locke's view of internal sense

⁴⁵ Kant seems not to have been aware that this requires a distinct set of concepts, i.e. psychological concepts.

was not specifically a theory of consciousness; for him both the internal and external senses are conscious. Armstrong used the view of internal sense to differentiate between two forms of consciousness, “minimal” and “introspective” consciousness, of which the latter corresponds to something that better deserves the label “consciousness” (Armstrong 1968). An absent-minded truck-driver may drive considerable distances on an “automatic pilot” before “coming to”, alerted by a sudden event, and so regaining introspective consciousness of his operations. Before coming to, the truck-driver must have had some sort of minimal perceptual consciousness of his surroundings to negotiate his way in the highway traffic. Armstrong’s concepts are plagued with conceptual problems, being inappropriate to differentiate higher-order from further first-order conceptual consciousness, and non-conscious from peripheral or even phenomenal consciousness. My interpretation is that introspective consciousness denotes some kind of reflective or access consciousness whereas minimal consciousness is more like (peripheral) phenomenal consciousness.

Of the most recent advocates of the inner sense view, William Lycan has forcefully defended Armstrong’s ideas. Leaning on Armstrong’s view, he characterizes the internal sense as a capacity to monitor the contents of our minds, as “the functioning of internal *attention mechanisms* directed at lower-order psychological states and events” (Lycan 1996, 14). Oddly, this is almost everything he says about the nature of inner sense. The rest of his account is fragmentary, and difficult to compile into a systematic psychological theory. He does refer to inner scanners as functionally unified but not necessarily as existing singly (*ibid.*, 32; cf. Carruthers 2000, 307–313). Furthermore, they have to be orchestrated with other mental systems to produce a successful interaction with the environment (Lycan 1996), and they also have the function of “relaying and/or coordinating information about ongoing psychological events and processes” (Lycan 2004). However, these descriptions leave the reader in the dark about the nature of the putative psychological processes at stake. A functionalist like Lycan should attempt to show in what kind of neural mechanism these functions are realized at least in the case of conscious human beings, given the massive advances recently in the cognitive neurosciences.

Another aspect of the HOP view is its “purity”, to what extent it is only perceptual, or, in other words, whether there is any conceptual component to it (see, e.g., Rosenthal 2004). Not all theorists have addressed this aspect, and it is not always clear by which means the attentional mechanisms are controlled. Armstrong, at least, (1997, 726) explicitly concedes that HOP does not rest on the mere “given” but is theory-laden. It is probable that a pure version is held by none of the HOP theorists. Nevertheless, they think that mere conceptual capacities are not enough, and that some non-conceptual kind of attentional awareness mediates conceptual awareness.

An interesting approach to inner sense is offered by Barsalou (1999) and Prinz (2007), although in neither case are their accounts offered as theories of consciousness. Barsalou’s ideas deviate somewhat from the inner sense theory as such, so I will deal with it later, but Prinz offers his view explicitly as a non-

conceptual view. He denies the existence of phenomenal concepts, and suggests instead that we have attention mechanisms that enable us to attend to regions of perceptual space in a manner that deserves the label “mental pointing”. In this view, our phenomenal knowledge is non-conceptual in that one should be able to redeploy the same abilities over multiple occasions of awareness of phenomenal states. This does not result in conceptual phenomenal knowledge for Prinz since he thinks concepts are something that can be recruited by memory in the absence of perception, while these attention mechanisms need the co-presence of perception. Others think that phenomenal knowledge is novel in every instance so that no general ability can explain its occurrence; phenomenal concepts are recognitional concepts (e.g., Loar 1997; McDowell 1994a; Carruthers 2000). Put this way, proponents of recognitional phenomenal concepts render them unable to fulfill the criteria for concepts in the terms of Prinz. Prinz’s view does admit of phenomenal concepts but not as direct or pure recognitional concepts. Phenomenal concepts can be formed to apply to experience only through mental pointing.

In my view, this is a problem for Prinz, since there is nothing that explains mental pointing to be pointing to the mental realm instead of being only world-directed. Mental pointing to redness can be directed either towards mental redness or towards worldly redness. Nothing in Prinz’s suggestion guarantees that pointing is directed towards *mental* redness. This is easily explained from a conceptualist view: redness is represented by psychological or phenomenal concepts as being mental whereas in the other case redness is represented by sensory concepts. Prinz could save his view if he could show that mental redness is originally given somehow as mental, or that we have some kind of non-conceptual way to discriminate between mind and world. Prinz does not address this issue, and no such account is easily defended, however, as will be shown in later in this chapter under the rubric of “psychic equivalence”.

The HOP theory is most vividly challenged by the other basic higher-order representational theory. According to the higher-order *thought* (HOT) theory⁴⁶, our higher-order representations are not perceptual but rather conceptual in nature and constructed not by a perceptual organ but by the conceptual capacities of thinking. The first critique of the HOP theory is that there have to be some sort of conceptual capacities at play in HOPs. How else could HOPs yield metacognitive understanding? The HOP theory is then more a theory about the privileged *access* that one has to one’s own consciousness. If, notwithstanding, the full picture of metacognition requires conceptual capacities as well, then the relation between higher-order perceptual and cognitive abilities has to be spelled out. HOP theorists should also better describe which sort of attention they are appealing to, voluntary or involuntary,

⁴⁶ The modern (actualist) version of this theory is developed in detail and defended by David Rosenthal in an important and influential series of papers (see, e.g., Rosenthal 1986, 1986; 1991; 1993a; 1993b; 1993c; 1997a; 1997b; 1998a; 1998b; 2000a; 2000b; 2002a; 2002b, 2002c; 2002d; 2003; 2004; 2006; 2007, of which many are collected to Rosenthal 2005).

spatial or non-spatial. To my knowledge, HOP theorists have not been able to respond to this requirement yet. Therefore, the HOT theorists point out, this problem, along with many of those mentioned above, can be avoided if one simply dispenses with the perceptual element; a satisfactory picture can be drawn using the conceptual capacities alone (Carruthers 2000).

According to the HOT theory, we are conscious of our first-order mental states simply by thinking about them, and thereby render them conscious. The picture becomes a bit more complicated when one notes that not just any kind of thought will do. There are two criteria for the thought to be a suitable thought. First, the thought must be directly about the first order sensation of red *as* a sensation of red (Rosenthal 2004, 4)⁴⁷. Otherwise, it would be about something else. This further criterion is supported by the function for higher-order states suggested by Rosenthal. We need higher-order states, since otherwise we could not become aware of perceptual *errors* (Rosenthal 2006) or adjust our beliefs about our mental states to each other (Rosenthal 2008). Some way to conceive of our mind being right and wrong is needed to cope in the world, given the errors made by our perceptual system, and this is why we have developed consciousness. Non-conscious creatures cannot make sense of their perceptual errors. (Rosenthal 2006 & 2007)

The account may at first sight seem circular: conscious thoughts are explained by other thoughts. It, however, need not be, since higher-order thoughts are seldom conscious themselves. Consciousness as the property of the state is the explanandum, and there is no circularity, since the state by which one is aware of the conscious state does not have that property. Put in the terms introduced earlier, state consciousness is *intransitive* and it is this kind of consciousness that we are trying to explain. We, however, can analyze it and explain it in terms of *transitive* awareness (for which there is nothing it is like to have that state). Consciousness is a relational property. The account is thus non-circular. Higher-order states can be conscious, and Rosenthal thinks that this is the case with introspection. Then, the second-order state is conscious, since a non-conscious third-order state is directed at it. A few more levels might be reached in the metacognitive hierarchy, but our cognitive capacity sets the limit at some point (Rosenthal 2000a).

One might, however, ask a follow-up question of how this squares with the alleged function of the higher-order state: how can such non-conscious states contribute to making sense of perceptual errors? Rosenthal might

⁴⁷ Sometimes Rosenthal has wavered with this. He says: "An organism cannot have HOTs about psychological states unless it has a concept of those states that characterizes them in the way relevant to HOTs. To have HOTs about thoughts and desires, the organism must have a concept of a state that has intentional content" (Rosenthal 2008, 837). Consider, however, the following citation: "For a HOT to attribute its target state to oneself, no more is needed than a conceptual distinction between oneself and everything else, a distinction that presumably any mammal can draw. Nor is any concept of mind required; HOTs *need not characterize their targets as mental*, but only as *states*." (Rosenthal 2000a, 207, emphasis PL, cf. Rosenthal 2005, 184). This, however, jeopardizes the theory as a *higher-order* view. How mental states can be distinguished from bodily states on such a view?

respond that they do so non-consciously. That much may be unproblematic, but then the function would be one of the non-conscious states, and the question of how consciousness helps in coping with perceptual errors remains unanswered. To my knowledge, Rosenthal has not dealt with this question yet, though he may have an answer to it⁴⁸.

Another aspect of the theory is that qualitiveness is already a state that first-order states have, and this can be analyzed in an externalist (Lycan 1996) or internalist way (Rosenthal 1991). In Rosenthal's example, one might have a non-conscious headache for a long time and then suddenly become aware of it, and the headache becomes a conscious headache; both states can have exactly the same qualitative properties, but there is something it is like to have that state only for the state one is aware of (Rosenthal 1991).

Moreover, the thought must be "roughly simultaneous". By this Rosenthal means, that the first-order thought and the higher-order thought must be present together in some way, and that the second must be the target of the first one. A thought that comes five minutes later cannot make its object conscious in the relevant sense anymore. A third and related criterion demands that the first-order thought must be available to the higher-order thought in a special way. In other words, the thought must have a non-inferential relation to its target. Even if it is a real possibility that one comes to a conclusion via inferential thinking and observations about one's own behavior that one must have a certain kind of thought occurring non-consciously, this inferential higher-order thought won't make the target thought conscious. What is a suitable way Rosenthal characterizes in the following manner: "A state will be conscious, on our intuitive understanding of that notion, if it simply *seems to one* that one is immediately or directly conscious of it" (Rosenthal 2006, 31).

One interesting issue in actualist higher-order theory is the relation between the lower-order state and the higher-order state. Rosenthal never explicitly describes the relation between them as causal. Rather, he describes it as an "accompaniment" and directedness that "results in" awareness of the lower-order state. This feature has naturally been balked at: some causal relation is seen as necessary, though it is not sufficient in itself for higher-order theories (Francescotti 1995). It is difficult to see why there should be any effects on the lower-order state if the relation is not causal. Since Salmon (1984), most postreductive views of explanation depend on some kind of causality, and thus this absence of causality it also deprives the theory some of its explanatory value. The relation between sensory and mental qualities being causal in his homomorphism view, it is interesting that Rosenthal refrains from causality in the case of higher-order states.

The HOT account is divided between two main accounts, the *actualist* and *dispositionalist* versions. The actualist version is the one advocated by Rosenthal. He takes consciousness to be something that requires actual occurrent HOTs, as it was briefly described above. Keeping in mind the distinction between

⁴⁸ At the time of writing this, Rosenthal is preparing a book on the function of consciousness.

occurrent and dispositional mental states, only a disposition to one is needed for consciousness. Conscious mental states are available to such higher-order dispositional states to become consciousness. Without suitable availability, mental states remain non-conscious. Note that this view is only half-way higher-order: it resembles the world-directed first-order theory in other respects than the possibility of higher-order representation.

The dispositionalist view has been defended in most detail by Carruthers (Carruthers 1996a; 1998a; 1998b & 2000). In its most radical form, the actualist HOT theory requires a separate HOT for every conscious element in consciousness. Aware of this threat, Rosenthal formulates his actualist thesis in such a manner that an indexical HOT will suffice for a range of elements in a given conscious mental state, thereby relaxing the stress imposed on a specific HOT. Nevertheless, Carruthers maintains that this picture is implausible in the sense that for every conscious state there has to be another state backing it up. The cognitive processing load is at least doubled. Appealing to indexical thoughts does not help Rosenthal; the actualist HOT model of consciousness remains vulnerable to the problem of *cognitive overload*. Consciousness requires an implausible amount of cognitive capacities. Such an amount runs also counter to everything that is known about brain functioning. A model of consciousness that could avoid the problem of cognitive overload will fare better.

A model of consciousness that avoids the problem of cognitive overload but still retains the benefits of the HOT theory and one which does not require a separate actual HOT behind each conscious mental state, is a dispositionalist HOT-model. On the dispositionalist version, no HOT is needed for each conscious state but instead only a possibility for the occurrence of the HOT. In the case of non-conscious mental states, there is no such possibility. The appearance-reality distinction can be accounted for in the dispositionalist model by insisting that the HOT system makes the distinction when it is actualized. This path is not, however, taken by Carruthers. Instead, he heads for a more minimalist view that appeals to a form of consumer semantics (Millikan 1984). In consumer semantics, meaning is determined in part by how the representational contents are used by the semantic system. On the view favored by Carruthers, one such consumer system is a "theory of mind module" that applies its ability to make the appearance-reality distinction to the original content: the perception of red acquires both the content "red" and "looks red" (Carruthers 2000, 242). This way the contents can gain the appearance-reality distinction even without the actual occurrence of HOTs on the conscious contents, and thus in a sense implement the idea of the transitivity principle. More precisely, consumer semantics explain how the conscious contents acquire subjective perspectivalness and their what-it-is-likeness character; by being available to the HOT/ToM (theory of mind) module the contents acquire also their subjective *feel*. Otherwise, mental states would remain without phenomenality; there would be nothing it is like for the subject to undergo such mental states. Their acquiring the content of subjectivity and phenomenality by being available to the HOT/ToM system furnishes them with these qualities.

Needless to say, Rosenthal does not agree insofar as he thinks that “a state is conscious only if one is conscious of that state, and being disposed to be conscious of something does not make one conscious of it” (Rosenthal 2000b, 8). Dispositional theories do not thus implement TP in the relevant way. Later, I will suggest that a dispositional theory can explain the intuition supporting TP, and thus works against it.

Carruthers is more specific about the status of a suitable explanation than Rosenthal. He points out that a naturalistic explanation does not require a classical reduction with identity relations but a reductive explanation that points out how the multiply realizable phenomena can be realized physically. In fact, even this is not necessary. For Carruthers, it suffices to render the account in a manner that describes the causal laws that the phenomenon obeys. Carruthers builds this picture in the form of theory theory functionalism where the functions are specified in relation to the whole of the system. Carruthers proposes that higher-order thoughts have the effect on first-order states of rendering them phenomenally felt by the subject. His model, Carruthers maintains, can explain the problem of phenomenal consciousness in naturalistic terms: it makes it understandable why with this kind of architecture there is something it is like to undergo some states but not others. A zombie with this cognitive architecture would be inconceivable.

Explanations need to show why something would happen or not happen, by appeal to causal laws. Such a causal explanation comes close to a causal mechanical explanation with the exception that in the mechanistic account, no laws need be invoked. Although I think his explanatory strategy comes close to the one I favor, I think that his overall theory fails: I expressed my doubts about his idea that knowing something would change its object already in the previous section.

As the dispositionalist view appeals to availability, it bears some resemblance to the GWS model of consciousness discussed in section 2.4. There are two important qualifications to this. First, a dispositionalist HOT view might only require availability to HOTs instead of global availability. Global availability, in contrast, does not necessarily involve a HOT system as one of its subsystems, so it is essentially a first-order theory. A version that would require the presence of HOT-systems would naturally constitute a (global) form of a dispositionalist HOT view. It should be noted that for Carruthers the HOT system works with atomistic recognitional concepts with no relations to other concepts. This raises the question about introspection: phenomenal knowledge could not be used at all. His view is compatible with some properties of phenomenal experience like ineffability. His view does not perhaps need to be so strict about the role of phenomenal concepts to guarantee their typical properties. This seems also to contradict his idea that the HOT system has evolved to enable cheating conspecifics.

Opinions for the “suitability” of HOTs diverge in Rosenthal and Carruthers. Rosenthal is more liberal about the suitability of HOTs. For him, any “assertoric attitude” that is directed to the first-order thought will do (Rosenthal 2005). Let us consider the example of wine tasting in more detail. A

novice degustator might first be able to discern only the coarsest properties of the wine, its acidity, sweetness, richness with some fruity flavours. With practice and careful analysis of its taste, the novice learns to make finer distinctions and becomes aware of more qualitative subtleties in the wine. She might start to differentiate in the acidic taste some softer and harder tannins and their balance, perhaps start to find more detailed elements of the bouquet of wines. Rosenthal suggests that it is by virtue of higher-order thoughts directed towards the wine-tasting experience that more of qualities of that experience become conscious. More specifically, the qualities must have been present in the non-conscious experience of the wine; it is implausible that the basic functioning of the sensory capacities would have been so different in the two experiences. The qualities in the experience are the same, and the difference is, on Rosenthal's view, when one becomes conscious of those qualities by suitable higher-order thoughts. Rosenthal points out that a qualitative mental state, the taste of wine, becomes conscious when HOTs are directed towards those states, and we become conscious of qualities that remained non-conscious before the occurrence of suitable HOTs to them. But, Carruthers has argued, the HOTs that make the new qualitative distinctions conscious are not directed towards the taste of the wine as a mental state but towards the taste of the wine as a property of the tasted wine. Rosenthal's view is based on his "homomorphism theory" according to which mental qualities resemble and differ from each other in the same way as the corresponding properties of physical objects. Rosenthal holds a kind of two-dimensional view of the content of qualitative states in that we become aware of the qualitative properties of physical objects partly by virtue of the qualities of our sensory states. This is guaranteed by a causal connection between physical and corresponding sensory properties; the object hence plays a part in determining the content. (Rosenthal 1991, 19–24.)

The difference between these two positions becomes clearer once we briefly remind ourselves of some of the basic features of the informational (representational) theory of intentionality in Dretske (1981). A signal carries information about many sources. For example, the ringing of a doorbell carries information about the doorbell but also about something that has caused the doorbell to ring, let alone all things that are nomically or analytically related to bell-ringing. In addition to the specific information that is carried by the signal, there is additional information "nested" in the signal. Dretske points out that there is, in fact, *no limit* to the amount of the information carried from a source. It all depends on the capacities of coding the information. What then specifies the actual source of the information in a given instance? The answer is the conceptual capacities that are used to extract the information. Similarly, in perception, information is carried not only about the source of information but also about the properties of the perception itself. In the wine-tasting example, these dimensions are the wine and the sensory perception of taste, respectively. Carruthers reminds that as long the object of the information is the actual worldly object, we move on the first-order representational level. Once the object is switched to the *perception* (a mental representation) of it we have a

genuine instance of a higher-order or meta-representation. Thus, Carruthers maintains, Rosenthal's account collapses into a first-order representational theory⁴⁹. Dretske's account gives at least reason to doubt the need for a space of mental qualities in addition to worldly ones, or the other way around.

If the object of a HOT cannot be the worldly object, but the mental state that is directed towards the object, adequate psychological coding capacities are needed to extract the specific psychological information. Not all theorists have acknowledged this point. Carruthers (2000) is the one that is most vividly aware of this, and later also Rosenthal (2005). For Carruthers, the answer lies in the debate in cognitive sciences and philosophy of mind over the *theory of mind* (ToM). Theories of theory of mind are usually discussed only in connection with the knowledge of *other* minds but seldom in connection with self-awareness, knowledge of one's *own* mind, Carruthers himself being a prominent exception (Carruthers 1997; Carruthers 2000). This is a crucial point especially in evaluating theories implementing the transitivity principle and their ways of explaining the possibility of the being aware of minds.

There are two basic kinds of theories of the knowledge of the psychological states of others and oneself. These are the theory theory and the simulation theory. The two have, in practice, exhausted the field with a couple of recent exceptions (e.g. Gallagher 2005, Bermúdez 2003). I will come to these exceptions once I have briefly described the main contenders.

The theory theory states that our understanding of the actions and psychological states of others is based on a conceptual framework consisting of our (folk-)psychological concepts. There are different versions of the theory depending how explicit or implicit this conceptual framework is seen to be. Views also differ as to whether this conceptual framework is a real theory in the sense of a scientific theory, or an inflexible theory-like construction that is genetically determined, as in the modularist view of Baron-Cohen (1995) The basic idea is in all forms that certain main folk-psychological concepts, such as belief, desire, memory, love, anger etc., inform a systematic whole that can be used in interpreting others and their behaviour. The most radical version claims that in the case of understanding oneself the system is radically the same application of the theory as in understanding others (Gopnik 1993). On this view, then, there is no real difference between first-, second- or third-person access to mental states. In addition to this behaviorist view, there exists a view, on which the access differs between first-person and second- or third-person cases, but there can be no real interpretation without the aid of conceptual systems (Carruthers 1997).

On the standard theory theory, understanding others depends on a person-neutral conceptual framework. The simulation theory claims to do without such an abstract mechanism. On simulationist accounts, understanding others comes about by running simulations of the perspectives of other persons on one's own first-person perspective (see papers in Davis & Stone 1995a and 1995b). One literally imagines oneself in another's shoes, and lets one's first-

⁴⁹ This point is elaborated extensively also in Aydede & Güzeldere (2005).

person perspective act as a *model* for what one would feel or do in that situation. This is then projected to the other person. The simulation is run more or less automatically, and will perhaps yield some feelings or intentions on the basis of which the other person is interpreted. Instead of person-neutral concepts, the knowledge base is one's own mind functioning as a model. There is one important aspect underlying simulation that is worth noting in the context of this study: in the simulation theory, the access to one's own mind is taken for granted. Theory theorists who have noted this have leveled the accusation that this sort of simulationism suffers from Cartesianism as far as the access is concerned: the access is direct, immediate and infallible (Carruthers 1996b).

There is also one definitely non-Cartesian simulationist account (Gordon 1995). Gordon claims that there is no such thing as an immediate access to oneself. Instead, knowledge of oneself can be gained by an "answer-check procedure" that works in the following manner. To know whether I have a belief, say about whether it will rain tomorrow, instead of engaging in introspection I only ask myself the question whether it will rain, and then check how I would answer. If in the affirmative, then I can conclude that I so believe. If the answer is negative, then I cannot veridically say the same. Carruthers has rightly concluded from this that the simulation account cannot provide a reasonable account of self-knowledge, it either slips into Cartesianism or behaviorism. Therefore, the simulation account cannot accommodate higher-order thoughts on their own terms (*ibid.*).

Another problem for the simulation theory is similar to a corresponding one facing the HOP theory: how can it claim to yield any sort of knowledge, or allow the retaining and manipulating of information about other minds? Simulationists are thus silent about the nature of both the nature of the privileged access related to and the conceptual abilities behind psychological self-awareness that were counted as criteria for a theory about it.

Simulationism might be a viable account but only as a hybrid theory with the theory theory, as the two do not necessarily rule each other out (Heal 1995; Carruthers & Botterill 1997). Such a hybrid view can, of course, avoid some of the problems of the less complex views by appealing to the virtues of whichever view depending on the issue. The role of the simulation mechanism is in this view complementary in the sense that simulations on the perspective of the other can be used as a heuristic device in interpreting the other. Otherwise we are supposed to do just fine with a "theory". Another, a more plausible hybrid view would be one where a process-simulation on the perception of others is run automatically to yield the raw data, and a theory is deployed when the finer interpretations are made. This view might also fall under the heading of simulationism, since not all of them deny the existence of a theoretical conceptual framework, despite claiming, nevertheless, that in everyday understanding of others the theory is not necessarily exerted.

The contenders come out evenly when related to some of the empirical data that pertain to this issue. The first and foremost evidence derives from experiments using the so-called false-belief (see Wimmer & Perner 1983) and appearance-reality paradigms (see Flavell 1986). These experiments have

suggested that children attain an explicit and proper understanding of minds at approximately three to four years of age. Before that they tend to claim that a sponge looking like a rock looks like a sponge to themselves and other people despite holding the opposite view just a moment before when not yet aware of the true nature of the object. Also they tend to predict the behavior of others on the basis of how things are, not on the basis how they think they are, even when given the information that others think otherwise, that is, have a false belief. Theory theorists take this to mean that by this age they have the full concept of belief, which is essential to the theory of mind, and which makes success possible in the theory of mind tasks of this level. Simulationists have explained the result by reference to the maturing of appropriate simulation capacities by which one comes to believe what the other believes. One acquires the ability to put oneself in another's shoes at three to four years of age, which can explain the performance in false belief and appearance–reality tasks.

Some phenomenological theorists, however, have nicely pointed out how, in these experiments, the required ability to use psychological concepts or simulation is an explicit, detached and theoretical one, preceded by implicit and more practical skills (Gallagher 2005; Zahavi & Parnas, 2003). Thus, toddlers and infants might have the ability to be primitively aware of consciousness, although this ability is not manifested in demanding and abstract experimental settings. Gallagher points out that because of their set-up, such false-belief experiments suffer from a problem of validity. They purport to measure the ability of subjects to understand other people's actions in general. The experiment, however, is a third-person situation where the subject requires a theoretical, detached understanding of the concepts of belief. Such an abstract meaning may not be applicable for them in an observational circumstance, an unnatural situation for a child. Nevertheless, they may be able to use these concepts in everyday second-person practical situations. This kind of discrepancy between practical and theoretical contexts has, in fact, also been shown in experiments on the A–R distinction (e.g., Rice et al 1997).

On the basis of these observations, Gallagher suggests making a distinction between two different forms of theory of mind (Gallagher 2005, ch. 9). One is the theoretical conceptual framework by which people's actions are explained from the third-person point of view. Before that, however, we have a skill, not exactly a theory – and not even properly concerning mind – but a practical non-mentalistic understanding, exerted in everyday social bodily interactions. This skill antedates theoretical knowledge, and is its developmental prerequisite. The practical understanding is based on the embodied interactions and perceptions of the bodily presence of other people in them. There is considerable evidence of such abilities even in newborns. The existence of mirror-neurons and the neonatal imitation serve as such evidence^{50,51}

⁵⁰ This practical understanding is not, however, an instance of the simulation theory, which these phenomena have also been taken to support (Goldman & Gallese 1998). For simulation one's mind would be used as a model; this is not the case in mirror neurons, since their activation is *actor-neutral*. Gallagher does not, however, deny the

Bermúdez (2003b & 2005) suggests that folk psychology is broader than is usually thought. In addition to attributing mental states to others, we have ready-to-hand heuristics and automatic routines for perceiving and understanding the patterns of behaviour of others, and thus we only seldom need to actually attribute mental states to others. Adherents of this view, however, are silent on whether these heuristics require previous non-routine uses of traditional mental state attribution that subsequently become a routine. It seems possible to acquire these in theory theory way or by virtue of a simulation. It seems possible to acquire them even in the way Gallagher suggests, so the view of Bermúdez can be seen as in line with all of these views.⁵²

In my view, the most fruitful and comprehensive account along these lines of the development of the A-R distinction is given by Fonagy et al (2002). Like Gallagher, for these authors theoretical knowledge is preceded by implicit skills based on embodied practices. However, on the view of these authors, such implicit practical skills are also products of long development. They are acquired in the early social relationships between the child and its caregivers, which Fonagy et al. conceive in terms of attachment theories (Ainsworth et al. 1978; Bowlby 1969)⁵³.

The traditional accounts of our understanding of others, the simulation theory and the theory theory have thus gained promising contenders nowadays, and the contenders seem to fare better than the original views of how people start to understand each other. These, however, do not help in the least the theories that try to anchor consciousness in psychological self-awareness. These views deny the existence of conceptual psychological capacities in more primitive forms of consciousness, such as in animals and children under three years old. Thus they constitute an important psychological reason to doubt the higher-order theories that are based on them, if we believe that both we in our first developmental phases and our fellow mammals are conscious. HOTs require conceptual capacities, and thus only the standard theory-theory is really compatible with the HOT theories of consciousness. If the recent views of the development of our psychological capacities are valid, then the HOT theory is compelled to deny phenomenality from animals and small children.

role and existence of simulation in understanding others but the most primitive and immediate form of the theory of mind is our practical and bodily understanding and interaction with others.

⁵¹ How then does this pertain to the transitivity principle? It is difficult to say. Despite holding the interaction to be action-oriented and non-mentalistic, Gallagher adheres in the following chapter of his book to the basic tenets of Husserlian phenomenology (Gallagher 2005). Is the A-R distinction a feature that continues to be included in the most primitive forms of understanding mentality? The status of the transitivity principle and its relation to the interaction theory remains unbeknownst to me.

⁵² These accounts also raise the issue that there is difference at what stage understanding of a specific kind of mental state is acquired. Understanding of attention and intention are acquired before desire (Meltzoff 1999; Baron-Cohen 1995; Astington & Gopnik 1991), and belief is acquired later than desire. Fonagy et al. (2002) take affective states to be the first ones to develop.

⁵³ More on this issue is given in the following chapter.

4.3 Intrinsicism and relationalism

Another useful way to distinguish theories of consciousness that implement the transitivity principle is by seeking to answer the question whether a given awareness of the state is accomplished by a further act of consciousness or whether each state is somehow aware of itself. On the first, *relationalist* view, givenness is a relation between two states. A conscious state is a state that we are aware of by virtue of another mental state, be that state itself conscious or not. In contrast, in the *intrinsicist* higher-order theories, or sometimes called *self-representationalist* or *Neo-Brentanian* views of consciousness (see, e.g., Kriegel and Williford, 2006), such self-awareness is an *intrinsic property* of each conscious state. There are not two states with distinct contents but each conscious state has two-fold content: in addition to awareness of the intentional content, there is awareness of one's awareness of the intentional content.

Sometimes intrinsicist theories pretend to be "same-order" theories in order to avoid the problems of relationalist views (Kriegel 2003). In my view, this label is not warranted except for some rare exceptions (Lurz 2003). This is because these theories implement the transitivity principle but they have simply transformed self-awareness from a relation between two states to a relation between parts of one state. Sometimes intrinsicists have been distinguished from relationalists by being labelled "one-level theorists" as distinguished from "two-level theories" (e.g., Zahavi 2005). In my view, this is a warranted distinction, but one must be careful not to conflate one-level theories with proper first-order theories of consciousness. One-level theories are, nevertheless, theories of consciousness that are based on self-awareness of mental states, like two-level theories. Therefore, according to my categorization, both belong to higher-order theories (theories based on awareness of one's own mental states), while it is only first-order theories that attempt to get along without any reference to such self-awareness.

There are several versions of the intrinsicist view (Kriegel 2003). All the versions described by Kriegel share the basic structure that consciousness is a result of self-awareness of a single mental state. A genuine same-order or first-order theory defines consciousness in terms of awareness of things, not mental states. Any view clinging to the transitivity principle faces the challenge of explaining the nature and possibility of being aware of mental states as opposed to other things. This task is even more difficult for the intrinsicist view, insofar as conceptual knowledge, privileged access and relation between the thematic reflexive states and their objects are rendered mysterious. Most intrinsicist theories rely on boxological diagrams of intrastate structures but remain silent about the psychological nature and mechanisms underlying their speculations (save Kriegel 2005). Symptomatically, the most detailed attempts to explain these aspects draw on the same resources as relationalist theories plus additional binding mechanisms and downplay the differences between intrinsicism and relationalism (Kriegel 2004).

Individuation of the mental state is thus the major distinguishing criterion between relationalism and intrinsicism. In addition to binding mechanisms, intrinsicists appeal to content in individuation (Kriegel 2005), while for relationalists the criterion is the attitude toward the content. Since there is always an assertoric attitude in higher-order states, then they must be different states (Rosenthal 2006). I am not wholly convinced by Rosenthal's argument that different attitudes within one state cannot exist. Could not one intentionally lie and at the same time in some way believe what one is saying? Moreover, I think that attitudes form a part of the representational system: attitudes represent things as having some value, and these can be equally assessed for accuracy or satisfaction.

Another aspect of the mysteriousness of the psychological nature of the intrinsic self-awareness of mental states is that no corresponding distinction such as the one between the HOP and HOT theories is present in the intrinsicist view. Intrinsicists tend to describe their views neutrally as based on "same-order monitoring". It is not clear what monitoring means in the intrinsicist view, since every state should have its own monitoring and the same monitoring could not continue from one state to another. It is also hard to see what kind of monitoring role attention would have inside one mental state unlike in HOP-view where monitoring consists of attentional mechanisms that constitute their own state. Monitoring that does not even rely on attentional mechanisms such as ones described by HOP-theorists (Lycan 1996) becomes difficult to fathom: no such form of attention is described in the cognitive sciences. Controversy prevails in the intrinsicist view over whether there is a marginal awareness that accompanies each conscious state (Kriegel 2009) or whether such awareness is some even more mysterious kind of awareness that does not show itself in experience (Gennaro 1996).

As mentioned earlier, relationalists hold that only they have the possibility to *explain*: if one takes consciousness to be an intrinsic property of mental states, one deprives oneself of any possibility of explaining consciousness. Intrinsicists have resisted this conclusion (Kriegel 2003; Gennaro 2005). The more positive reasons for intrinsicism are phenomenological: it is a phenomenological feature of experience that our awareness of conscious states is an intrinsic property of the state, not a different state (Kriegel 2003; Kriegel 2009). Also the features that Rosenthal requires for access, simultaneousness, non-inferentiality and privilegedness fall more naturally from the intrinsicist view. A relationalist view could show in response that it can be formulated so that it explains this phenomenology. Rosenthal, however, thinks that the mental state by which we are conscious of our mental states does not have to be and normally is not conscious at all (Rosenthal 2005)⁵⁴. A similar move is naturally available to intrinsicists as well (Gennaro 2006), and a lively debate continues between these different views.

⁵⁴ This implies that there would be no trace of the transitivity principle at the level of phenomenology, which is contrary to the phenomenological point just made by intrinsicists. One might wonder why people would usually think that the transitivity principle holds if it is not manifested in phenomenology.

Moreover, it has been claimed that many phenomenological views of consciousness inspired by Husserl endorse a version of the intrinsicist view (e.g., Kriegel 2003, 2004). Indeed, there are elements in at least some phenomenological views, also descendants of the Brentanian theory, which bear a resemblance to the intrinsicist view (Smith 1986; Zahavi 2004; Thomasson 2000). To be precise, phenomenologists do not necessarily build theories of consciousness that explain consciousness. Rather, they only describe the essential (or typical) features of a phenomenon. They do, however, go behind primarily given appearances in that they purport to describe the conditions of possibility for the properties they are describing. When describing these properties and conditions, they tend to approach the intrinsicist view, and cast themselves as clinging to self-awareness.

On the Husserlian phenomenological view originating in the *Logical Investigations*, not only are objects given to the conscious perception (*Wahrnehmen*) but the act by which we come to be aware of objects is consciously experienced or lived through (*Erleben*), and this is what makes object-givenness possible in the first place. Living through the act is a necessary condition for being conscious of objects. Dan Zahavi interprets this in such a manner that living through the act is a necessary condition not only for intentionality but also for qualitative feel, the hallmark of conscious experience. This much perhaps is uncontroversial among the phenomenological theories of consciousness. Zahavi then proceeds to make the strong interpretation that to live consciously through the experiential act means to be pre-reflectively *aware of the experience*⁵⁵. Consider the following citation from Zahavi:

“To undergo an experience necessarily means that there is something ‘it is like’ for the subject to have that experience. But, insofar as there is something ‘it is like’ for the subject to have experiences, there must be *some awareness of the experiences*. In short, there must be some minimal form of self-awareness. To be acquainted with an experience in this first-personal mode of givenness is to be in possession of a primitive type of self-awareness, and, on this account, the only type of experience that would lack self-awareness would be an experience the subject was not conscious of, that is, an ‘unconscious experience.’” (Zahavi 2003, p. 88, emphasis PL)

In light of this citation, Zahavi seems to define consciousness in line with the transitivity principle. Consciousness thus involves a “pre-reflective self-awareness”, in which conscious experience is given to itself (cf., e.g. Zahavi 1998, 22–23). Zahavi is undoubtedly right to claim that later on Husserl explicitly makes a similar commitment in his analysis of inner time-consciousness. In the so-called retentive aspect of consciousness, there is twofold intentionality: the elapsing phase of the object is retained in consciousness but, by the same token, the previous *act of consciousness* is also retained⁵⁶. Husserl calls these transverse and longitudinal intentionality,

⁵⁵ I will deal with an alternative, weaker interpretation in the last section.

⁵⁶ Shaun Gallagher interestingly suggests a similar structure to protentive or anticipatory aspect of consciousness as well (Gallagher 2005, 194).

respectively. In Husserl, conscious experience includes awareness of itself in the form of longitudinal intentionality.

In addition to the possibilities of qualitative feel and for object-intentionality, there is yet another thing phenomenologists take to depend on this pre-reflective self-awareness⁵⁷. One can be aware of one's mental states, if not pre-reflectively then at least reflectively and thematically simply by reflecting on them. A theory of consciousness must be able to account for the possibility for such reflective consciousness. For Zahavi, it is precisely the same givenness or pre-reflective self-awareness of consciousness in the form of longitudinal intentionality that guarantees the motivation and possibility for further thematized reflective acts (see Zahavi 1998, ch. 7).

There is, however, a decisive difference between phenomenological theories and the relationalist and intrinsicist higher-order theories insisted on by Zahavi. He makes the reservation that longitudinal intentionality should not be taken as intentionality of the subject-object form or as any sort of object-awareness (Zahavi 1998, ch. 2 and 3). Such views lead to a host of insurmountable problems of circularity and infinite regress (see section 4.5), and Zahavi has shown convincingly that Husserl was aware of this (*ibid.*). If, then, talk of transitivity is not quite warranted in the case of the phenomenological theories, they should not perhaps be categorized as theories implementing the transitivity principle or at least the transitivity part of it, with the other intrinsicist theories. Nevertheless, they share the basic idea espoused by the transitivity principle that there can be something it is like to undergo experiences only if one is in some, intransitive, pre-reflective, non-conceptual way aware of it. Thomasson has labeled this view the "reflexive content view", as opposed to a dual-state or dual-content view (2005). Following Smith (1989), she holds that experience is not endowed with two distinct contents but one complex content "I see consciously X", with the emphasis on X.

It may be hard to many to imagine what such an intransitive and non-conceptual awareness (even of experience) could be like. Whatever the answer to this question, the phenomenological view retains, however, one important feature that is common to the theories with explicit adherence to the transitivity principle. Were the awareness of the experience transitive or not, also on this phenomenological view consciousness *must involve the A-R distinction as an intrinsic property*. In order for experience to be given as experience, albeit in an intransitive manner, the A-R distinction must be in place. Otherwise, neither the givenness of object and act, nor reality and appearance could be distinguished in the experience.

⁵⁷ There are even more things such as the unification of consciousness (see, e.g., Husserl 1968, 43). I will not deal with these other issues here.

4.4 HOR and language

During the resurgence of consciousness studies in 1970s and 1980s it was popular to think that consciousness requires language. Fortunately, this view has been slowly dying out. In the context of the higher-order theories of consciousness, this view still possesses some ground. It can be claimed that metacognition makes one conscious and that metacognition is not possible without language. Daniel Dennett, makes this sort of contention specifically about consciousness, while others, e.g. José Luis Bermúdez, make it about the nature of metacognition.

Dennett's earlier theory (1978) falls neatly within the category of dispositional higher-order thought theories with the exception that for him, higher-order thoughts are linguistic. Dennett's theory has undergone some modifications. Some of the aspects modified in his later view are the following (Dennett 1991). Linguistic cognitive systems all the time produce competing interpretations of contents, "multiple drafts". This non-conscious competition operates in parallel. Some of the contents produced by this system are selected for broadcasting to various parts in the brain via the short-term memory. On top of the parallel processing is a quasi-serial linguistic system, the "Joycean Machine" which connects the winning contents manufactured by the parallel competition into a sensible connected stream. By this, Dennett ensures that he can accommodate three things. One is the facts we know about cerebral specialization. Second is the advantage of connectionist models of cognition where neural networks are responsible for the contents; in Dennett's case, there are many such networks operating in parallel. Thus, Dennett attempts to reconcile the modularist and connectionist views, which had not yet been formulated in the 1970s. The computationalist intuition that we process symbols serially is guaranteed by the Joycean Machine. Thus, Dennett accounts for our subjective view that the conscious stream of thought operates serially with distinct theories about cognition and neurophysiology. This kind of model avoids, according to Dennett, the model of a "Cartesian theatre" where consciousness is a single homunculus-like observer for whom the contents are made available. Further, Dennett reminds us that nothing we know about cognition or the brain supports this. Notwithstanding, the overall picture of dispositional higher-order thought applies to Dennett's multiple draft model: consciousness is "cerebral celebrity" where the competing "linguistic demons" formulate phrase-candidates of which one is selected to the overall stream of the contents of the short-term memory. These contents are globally available. In fact, the contents are not made available to a linguistic system but are already manufactured as linguistic before being sent to the short-term memory. The overall model thus largely resembles the dispositional ones of Carruthers and Baars.

Dennett's has been widely criticized as an eliminativist or a behaviourist theory, but it has some virtues. In particular, I share his aversion to conceptual

pseudo-problems such as the problem of qualia (see, Dennett 1988 & 1991). Instead of chasing pseudo-problems we should look at what we know about consciousness and cognition and its neurophysiological basis. This is in line with philosophical naturalism.

It is interesting that the linguistic HOR view is also put forward by another proponent of philosophical naturalism, José Luis Bermúdez. In his HOR, or as he describes it, “intentional ascent”, view, for metarepresentational thought or intentional ascent to be possible, a (first-order) thought requires a *suitable vehicle* by which it can be held in mind so that a (higher-order) thought can be directed to it. Moreover, only language provides the kind of vehicle, which renders metarepresentational thought possible – “intentional ascent” requires “semantic ascent”.

As the only imaginable alternatives for a suitable vehicle, Bermúdez considers Johnson-Laird’s “mental models” account of thought and Braddon-Mitchell and Jackson’s (similar) mental maps theory. These theories take thought not to be language-like but rather imagistic in form; that is, isomorphic to what it represents. The problem, according to Bermúdez, concerns the thesis of the structuredness of the mental models that allows inferences and their evaluation: “[T]heir (mental models and mental maps) structure is derivative. It is derived from the premises that they are modelling. The models are constructed from constituents and properties that feature in the premise being modelled. And those premises are of course linguistic entities” (Bermúdez 2003, 163). Therefore, the mental models can connect to one another only in a sentential manner. Consequently, the mental models theory ultimately collapses into a special form of linguistic thinking. Bermúdez insists that the vehicles of metarepresentational thought in all their forms can be available only to language-using creatures because there is no non-linguistic alternative available as a vehicle for such thoughts.

I fail to see why imagistic thinking is derived from sentential or pictorial thinking composed of linguistic entities, as Bermúdez maintains, rather than the other way around. Moreover, I especially fail to see why metarepresentational thought that Bermúdez refers to here concerns only inferential thinking. Why would all second-order cognitive dynamics involve inferential relations? Leaving these questions aside, I think Bermúdez does not exhaust the alternatives. More precisely, Bermúdez concedes that there is metarepresentational thinking that has sensations and mental images as objects, but he does not consider sensations and mental images as instances of genuine thinking. He thinks that “we are not [...] ever conscious of propositional thoughts that do not have linguistic vehicles” (Bermúdez 2003, 160). Restricting the scope of thought this way, however, Bermúdez leaves no room for thinking that has no propositional structure involving, for example, only nominative and predicative components as in his own theory of non-linguistic thought. I will come back to this in more detail in section 4.

Bermúdez rightly points out that metarepresentational thought is not a monolithic phenomenon. He lists three forms of explicit intentional ascent. Two of them, second-order beliefs and second-order desires, are intrasubjective,

directed to one's own beliefs and desires. The third form of explicit intentional ascent, theory of mind (ToM), is intersubjective and necessary for understanding others. In addition to these explicit forms of intentional ascent, Bermúdez maintains there are several forms of implicit intentional ascent. All of them involve modes of complex compound thoughts where entire thoughts are objects of further thoughts. Thoughts of this kind include not only truth functional thoughts ("it is true that..." / "it is false that...") but also involve modal, adverbial and tensed thoughts. All of these take one thought as the object of further thoughts. It is possible to counter-argue that these are not forms of metarepresentational thought because they concern the states of affairs rather than the thoughts themselves. But Bermúdez replies that, in addition to being about states of affairs, the thoughts must involve an implicit form of higher-order thinking for the evaluation of their content to be possible. Unless the thought is first taken as a truth-bearer such as thought, we could not evaluate whether it is true or false.

All this has the obvious implication that creatures that have no linguistic capacities are incapable of metarepresentational thought. These creatures include most mammals and birds, and the members of our species deprived of linguistic capacities, such as infants and aphasics.

Bermúdez puts his argument forward as a transcendental argument: in the absence of any other kind of vehicle for meta-representational thought, such thinking is available only to linguistic creatures. But, in the general spirit of the philosophical naturalism advocated by Bermúdez, the argument that intentional ascent requires semantic ascent should also reckon with the empirical data that is available. Bermúdez does adduce some empirical data; he deals with the debate over the existence of ToM in non-human apes and monkeys (baboons) and concedes that the general opinion favors the supporters of ToM in apes and baboons. Nevertheless, he adheres, in a spirit akin to Morgan's Canon, to a more parsimonious interpretation of this data, according to which the existence of ToM in non-linguistic animals is dubious and the data explicable by mere abilities of detecting perceptual states such as gaze-direction.

It is true that experiments with apes and baboons have not succeeded in settling the issue definitively. However, there are data bearing on the issue that he seems to overlook. First, there are experiments that suggest infants have the ability to recognize the intentions of others at the age of no more than 14 months (Meltzoff et al, 1995; Call et al, 2004). If infants see adults failing in a simple task, the infants imitate the actions not as they were performed but as they were clearly intended to be performed. Second, there is a recently reviewed article on uncertainty monitoring that suggests that non-linguistic animals possess some form of metarepresentational thought (Smith et al. 2003). Smith et al. concede that it might seem more parsimonious to interpret their data in non-metarepresentational terms, but they argue that this is not, in fact, the case. The reason for this is that they are compelled to postulate different kinds of mechanisms underlying the performances of humans and animals even when those performances are very similar. One could say that this is a

beginning of a metacognitive turn, and thus a return to the Morgan's Canon is not possible in the case of metacognition, any more than in the case of cognition in general.

One must concede that these data are suggestive rather than conclusive. But there seem to be yet further empirical data that are more compelling. Peter Carruthers changed his mind (from his work of 1996 to that of 2000a) about whether higher-order thought is language-bound on the basis of data provided by Rosemary Varley's (1998; 2001) experiments with a-grammatic aphasic patients: patient SA passed a non-linguistic version of false-belief test administered by Varley. The evidence on aphasics seems, eventually, compelling enough to question the slogan "intentional ascent requires semantic ascent". The fact that the first data were obtained using notes containing single words hints at the use of language. Nevertheless, this cannot be language in the sense of Bermúdez as for him language requires a sentential structure. Rather, it appears to be an example of thinking with a nominative component and a predicative component - an example that Bermúdez himself gives of non-linguistic thought. It seems that this level of thought is enough for being capable of forming judgments about other peoples' beliefs.

Nevertheless, the evidence from aphasia for non-linguistic metarepresentational thinking leaves us with the following problem: What are the vehicles of such thoughts? Let me consider one possibility that I think might be a plausible alternative

The theory of Bermúdez (2003) is, in fact, not a theory of consciousness *per se*. It is more an exploration of the limits of non-linguistic thought. Bermúdez claims, in a form of a transcendental argument, that the limit of non-linguistic thought is in metacognition. Metacognition is not possible without language. For higher-order theories of consciousness, the most important implication of Bermúdez's views is that higher-order representations must be conceptual (thinking, not perception). Higher-order representations cannot be but linguistic. Thus, if consciousness requires HOTs, a conscious creature must master a language. Bermúdez arrives at this conclusion by a transcendental argument. This is odd considering his longstanding criticisms of transcendental arguments that they should be sensitive to empirical data. And here he does not consult the relevant data, even data which can be found in his references, i.e. data which point to metarepresentational capacities in non-linguistic creatures.

4.5 Appraisal: Insufficiency of HOR in explaining phenomenality

Higher-order theories have faced severe criticism on a number of fronts. Some of the problems of the specific theories were presented in connection with their presentation. Here, I shall chart and evaluate some of the arguments presented against HO theories more systematically. I have collected a comprehensive if not an exhaustive list of counter-arguments in Table 1. My purpose is to point out that there are two main problems from which more specific problems stem. I will argue that the first of these is crucial to the HO-theory as a theory of phenomenal consciousness, while the second is not.

TABLE 1 Objections to HO theories

General problems

1. Explanatory insufficiency (Hardcastle 2004)
2. Fallacy of representational divide (Güzeldere 1997)
3. Strandedness (van Gulick 2006)
4. Generality problem (Van Gulick 2006, Rey 1983)
5. Impossibility problem (Searle 1992; Dretske 1995)
6. Awareness without HORs (Dretske 1993)
7. Causality problem (Francescotti 1995)
8. Extra conditions problem (Van Gulick 2006)
9. Hard problem of consciousness (Stubenberg 1998; Siewert 1998)
10. Bad theory problem (Chrisley 2004)
11. Epiphenomenality (Dretske 1995, 117)
12. Misrepresentation (Neander 1998; Byrne 1997; Levine 2001)
13. False positives (Neander 1998, Byrne 1997; Levine 2001)
14. False negatives (Block 2002)
15. Empirical implausibility (Hardcastle 2004)
16. Cognitive overload (Carruthers 2000)
17. Attentional deficiency (more consciousness than we can attend to)
18. Cannot distinguish absent-minded and blindsight (Stubenberg 1998)
19. Animal/infant consciousness (Carruthers 1989; Seager 2004)

Self-awareness related problems

20. Fallibility (Shoemaker 1994)
21. Regress (Zahavi & Parnas 1998)
22. Circularity (Zahavi & Parnas 1998)
23. Immunity to error through misidentification (Baker 1998)
24. Non-naturalism (Baker 1998)

Relationalist HOT theory	Intrinsicalist HO-theory	Dispositionalist HO theory	HOP theory
1. Immediacy (Francescotti 1995)	1. HOP-HOT distinction	1. Effect problem (Rosenthal 2000b, 2004)	1. Sense organ (Sturgeon 2000)
2. Phenomenology (Kriegel 2004)	2. Mysteriousness (Rosenthal 2004)	2. Suspect semantics (Rosenthal 2006)	2. Sense qualities
			3. Complexity (Carruthers 2000)

I will deal briefly with each objection that has not been dealt with earlier. The first problem is that the whole enterprise seems to stand on dubious intuitions.

Problems 2 to 19 are more or less manifestations of the first problem. The second strand of problems, those from 20 through 24 are related to reflexivity of the relation between LO and HO states, the fact that they belong to the same subject. There are, furthermore, a number of problems with different versions of higher-order theories that stem from their specific structure. Most of these have already been dealt with in connection with their descriptions.

1) The most general and at the same time the most pressing problem, in my view, is that the transitivity principle is presupposed, not explained. I call this problem that of “explanatory insufficiency”. Save perhaps Block (1995), Hardcastle (2004; see also, Chalmers 1996; Siewert 1998; Stubenberg 1998) has in my view best identified the real core of this problem:

“At this point, the burden falls onto the HOT theorists to explain exactly what they think is going on here and why that results in consciousness. They now wading deep into empirical waters. If they don’t follow through at this point with ways to operationalize their account, then they are open to the charge that they are explaining the mysterious by the more mysterious. The ultimate aim of HOT theories is to explain conscious mental states in such a way that the mysteriousness of consciousness is removed (or at least diminished). But if they are telling us that consciousness results from two structurally similar mental states being co-active, then this obviously does not lessen any mysteries. It simply shifts them over from the qualia to the co-activation. Furthermore, it still leaves wide open the question of why it is this co-activation that causes or results in consciousness. [...] Why should a HOT confer consciousness, of any sort and at all? Without further development, we have no explanation. In virtue of what do HOTs make us conscious of anything? Without answering this question, then Rosenthal is either begging the question by trying to explain consciousness in terms of consciousness or is waving his hands at the problem and calling it solved.

As discussed above, the reason Rosenthal opts for a psychological account of consciousness is that he buys into a version of the explanatory gap: he can’t for the life of him see how anything neural could cause something as weird and as grand as a conscious experience. However, nothing that he has done so far reduces that very same claim against him. Why should two co-occurring mental states result in qualia? That is just as wild and weird as neuronal interactions giving rise to phenomenology.” (Hardcastle 2004, 287-288)

The insufficiency alluded to is aggravated by the fact that there are intuitions to the contrary that are equally compelling, so a more careful analysis may be needed to settle this question. Opponents of the transitivity principle can equally appeal to everyday intuition. Consider a primitive form of consciousness, e.g., that of a rabbit perceiving an approaching fox. For obvious evolutionary reasons, even a brief glimpse of the fox immediately captures the attention of the rabbit, and the fox and the urge to flee from it fill the rabbit’s consciousness. It sounds highly exaggerated to maintain, as a higher-order theory would, that, under such circumstances, the rabbit’s elevated consciousness is due to its being in some way aware of its own mental states, thereby making them conscious. On the contrary, it is generally the case that consciousness of one’s mental states would *weaken* one’s consciousness of outward objects.

The argument is not, perhaps, restricted to animals. There are phases of human development which seem reasonable to see as equivalent of animal consciousness in this respect. It might not even be restricted to earlier phases of the development of human consciousness. Such phenomena as mental absorption, flow experience, ecstasy or trance may be cases in point. These states are allegedly characterized by one's occupation with the object to the detriment of one's awareness of oneself, let alone one's mental states. Perhaps the point made about the rabbit's natural terror of foxes applies equally to corresponding adult human terror. If one becomes instantaneously terrified of something threatening, e.g. when coming upon a growling bear on a walk in a forest (as once happened to me), it seems exaggerated to maintain that one's elevated consciousness is due to one's elevated awareness of one's consciousness rather than the object and its immediate relevance to one's well-being.

Rosenthal might have the resources to respond to this, since he described as relevant higher-order thoughts those that subjectively seem immediate to us. Other mental states, such as conscious ones do weaken the consciousness of their objects. However, this response remains as unexplained as the general idea. The important question is: Why would being conscious of something make its object conscious? Mere suppositions or convictions are not enough. Faced with these conflicting intuitions, my intuitions lie with those of the opponents of the transitivity principle. In my view, there is reason enough to be suspicious about the transitivity principle. It cannot only be taken for granted. It is at least fair to ask: to what extent does it really hold? If it turns out not to hold or no explanation for TP itself cannot be given, it would mean that the big bubble into which HO theories have been pumped is full of explanatorily thin air as far as the phenomenality of conscious experience is concerned. This possibility calls for further analyses.

Rosenthal has defended the transitivity principle by presenting examples where he argues that the transitivity principle is implemented, like that of wine tasting. However, we saw earlier that the wine-tasting example is invalid. Learning new words to describe something, e.g. tastes of wine, contributes not to the higher-order thoughts about the mental states that are about those objects, but first-order states about the object itself, *wine*. Higher-order thoughts must be about the mental states to be higher-order, not the objects the mental states are about, although the concepts classifying the properties of objects may contribute to the development of the phenomenal concepts (concepts about the phenomenal properties of mental states) from the sensory concepts (concepts about the sensibly perceivable properties of objects) (see, Aydede & Güzeldere, 2005). Why, then, should a higher-order thought boost our consciousness of the first-order content as more vivid, if it is not about that content in the first place? If it is the case that learning something about the wine we taste makes our conscious mental states more vivid, what is the reason for this? It may be more plausible to argue that not being conscious of our mental states but rather more vivid consciousness of *objects* is what makes a state conscious.

The line of thought that cashes out consciousness precisely in this way, thus denying the transitivity principle any role in explaining consciousness, is first-order representationalism. Dretske insists that consciousness is not awareness of conscious acts but rather consciousness of *objects*: “Conscious mental states – experiences, in particular – are states that we are conscious *with*, not states we are conscious *of*” (Dretske 1995, 100). Consciousness is characterized by its *transparency* or *diaphanousness*: Primordially, what is given in the experience are not mental properties but properties of things in the world. Put in more phenomenological terms, consciousness always has a *self-transcending* character (Zahavi 2005, 282–283).

2) The “fallacy of the representational divide” coined by Güzeldere (1997) makes essentially the same point by focusing on the concept of representation. In HO theories, the force of the HO state should yield not only awareness of the LO state but also more vivid awareness of the outer object. This, argues Güzeldere, is not what HO states are equipped to do. Consider once more the example of wine-tasting. In order for one to become conscious of various aspects of the wine, it is not enough that one becomes more conscious of one’s tasting but one should become able to discern more accurately and stably the properties of the wine. The force of the higher-order state does not carry over to the other side of the representational divide. Güzeldere asks: “Why try to explain that [externally directed awareness, PL] in terms of another form of, *internally directed*, awareness” (Güzeldere 1997, 796). A first-order view that takes further states to target the same first-order content with new conceptual capacities naturally has no problem with this.

3) Van Gulick (2006) makes a similar point again about phenomenality: the aspect of internal directedness is not phenomenologically present in the conscious state at all. Consciousness is “stranded” from the first-order state. This can be taken to support intrinsicism, like that of Van Gulick, but a similar, though fainter threat of strandedness remains for intrinsicism when compared to first-order views.

4) Despite the name, the “generality problem” is a particular aspect of the dubiousness of the transitivity principle (Rey 1983; Dretske 1995, Van Gulick 2006). Many have questioned why the transitivity principle does not hold generally, e.g. being aware of one’s liver does not make one’s liver conscious. The obvious response given by the HO-theorist is to restrict it to hold only for intentional (Rosenthal 1997b) or psychological states (Gennaro 2005). This of course only begs the more general question of explanatory insufficiency: why does the transitivity principle only hold for these phenomena and not others? Unfortunately, the discussion has not properly reached this more general question of why it should even for psychological states.

5) Dretske has challenged the HOR theories on a number of fronts in addition to the generality problem. One objection stems from his informational account of consciousness. On his account, informational systems carry information about distal sources; this being so the informational system cannot at the same time be the source of the information. A signal may carry information about the informational system as well but the informational

system cannot be the primary source of the information-carrying signal. The signal must be from a distal source, but depending on the encoding system, the signal can be encoded in a manner which might indirectly use the information about the informational system that is nested in the signal. In other words, we cannot be directly aware of our representations. Introspection is not something that is directed towards the representations themselves but to worldly objects. What is needed further is a suitable belief in which the information about the informational relation itself is digitally contained. In introspection, one cannot observe one's mind as such, but the outer reality. We can only be directly aware of representational contents, not the representational vehicles, which presumably consist in populations of neurons to which we have no direct access (Dretske 1995; cf. Searle 1992). In other words, no direct higher-order representation can exist on which HO theories are based in the first place. I label this problem as the "impossibility problem".

As a follow-up to the impossibility problem, it would be natural to see Dretske's theory of introspection as superior to the HO theories. However, its problems should be born in mind (see section 4.1). These apart, showing the problem, that the objects of introspection are something other than representational contents, can undermine the impossibility problem. It is possible to think that there can be higher-order representations and yet hold that it is a problematic theory of phenomenal consciousness. Once a first-order representationalist conceded this, she would not perhaps necessarily embrace a HO theory of phenomenal consciousness but she could embrace a HO theory of introspection. In fact, I think this is the path that should be taken, and I will attempt to take it in the next chapter.

6) The second and related argument by Dretske is that there are aspects of the contents of first-order states that remain non-conscious despite our awareness of the first-order state. This is well illustrated by the phenomenon of change blindness⁵⁸. It has been shown that a significant change in two otherwise identical images often goes unnoticed because of some transient secondary distracting stimulus or event, such as a brief blank screen or eye-blink. The changes are such that without the distraction they would be detected immediately and, once the subject becomes aware of it, the change becomes eye-catching and it is hard to believe that such changes, like the moving of a building or a change in the identity of a person one is having a nice chat with, go unnoticed (Simons & Levin 1997; Rensink et al. 1997). Dretske asserts that here the perception of these images or views containing changes is clearly

⁵⁸ Change blindness has been brought up also as counter-evidence for representationalism in general. Instead of containing detailed sensory representations of the environment, consciousness is constituted by crude knowledge of how to act in relation to one's environment ("sensory-motor contingencies", O'Regan & Noë 2001; Noë 2004). Change blindness is supposedly due to the sparseness of our representations: the changed features are not represented and changes become noticed only when focal attention is directed at them. However, there is a wealth of experimental data showing that changes are implicitly noticed, and that they affect performance and neural measures (Fernandez-Duque & Thompson 2000, 2003; Hayhoe et al. 1998; Lyyra et al. 2010).

conscious: one sees the changes but does not *recognize* them or is unable cognitively to respond to them. What a HOT theorist is compelled to say is that there is a HOT that makes one's perception conscious. However, Dretske adds, it does not succeed in serving its purpose of making one conscious of the change. Therefore, HOTs are insufficient in explaining consciousness (Dretske 2004; cf. Dretske 1993). This counter-argument is akin to that of false positives (13), but it differs from it in that it attempts to show that if there were HOTs, they could not explain consciousness. The false positive counter-argument is that there are in fact suitable HOTs but they fail to make target states phenomenally conscious.

7) The causality problem concerns the relation between HO and LO states. Francescotti (1995) is the most vociferous critic of HO theories in this issue. He insists that the only way to make sense of the relation and the force of the transitivity principle is to appeal to causality rather than mere accompaniment, as Rosenthal does. No account of causality, however, can be accommodated without problems into the HO theory. Causality and causation has been a tricky subject about which I have not much to say (for interesting discussions on the metaphysics and explanatory issues on causality, see, e.g. Woodward [2003] and Craver [2007]).

8) The "extra conditions problem" pointed out by Van Gulick (2006) is one that follows from the response to the generality problem. In order to guard itself from problems like the generality problem, an HO theory has to provide additional conditions for its validity. Although they might keep the HO-theory standing, it no longer looks like a monolith standing alone but is rather a shaky edifice supported by various minor ones.

9) Against this background of explanatory insufficiency and related problems, it comes as no surprise that it has been complained that HO theories do not, in fact, address the hard problem of consciousness (e.g., Stubenberg 1998, Siewert 1998). What may come as a surprise to many is that this is not, in fact, denied by the HO-theorists (Rosenthal 2005, 190-194; Lycan 1996, 77; cf. Carruthers 2000, 127-128). The hard problem of consciousness deals with the concept of phenomenal consciousness. Of these prominent HO-theorists Rosenthal and Lycan (the view Carruthers on this issue is not clear) have expressed their dissatisfaction with Block's initial formulation of the distinction between phenomenal and access consciousness. It is symptomatic that they claim not to be dealing with mere phenomenal or qualitative aspects of consciousness, but hold that consciousness, there being something it is like to be in those states, amounts to something more, namely, an awareness of those states. This aspect does not necessarily render the HO theorists any weaker, but rather makes them stronger. Their refusal to explain phenomenality as such in practice deflates most of the counter-arguments leveled at it. Thus the HO theories purport to explain some other form of consciousness, and may be entirely up to the task. Then, the obvious question is: which form of consciousness are they explaining? The issue remains unclear. I shall give my own account in the next chapter where I claim that HO theories can describe the

common kind of human consciousness that I shall call “mentalized consciousness”.

In contrast to most of the HO theorists, I believe that the concept of phenomenal consciousness does refer to a genuine aspect of consciousness. However, I agree with the HO theories that an important aspect of what it is like to be conscious in the way human beings are, consists in being at least potentially aware and in control of one’s mental states. However, this aspect of consciousness is something above mere phenomenality, and I think the concept of mentalized consciousness captures the intuition supporting the HO theories. By the same token, many problems are avoided.

10) HO theory is also accused of being too strict on creatures with a “bad theory of mind”, insofar as it is compelled to deny that they are conscious (Chrisley 2004). This observation aggravates the doubts that higher-order representations may not be responsible for consciousness.

11) Dretske also complains that the idea that higher-order representations are the source of consciousness make consciousness epiphenomenal since HORs do not add to the causal powers of the target mental states (Dretske 1995, 117). This seems to drive deeper the wedge between consciousness and higher-order states. Rosenthal may try to respond to this by asserting that by error detection the causal powers of a conscious creature himself change, if not the causal powers of the first-order state. In this case, self-awareness makes a difference in the form of creature consciousness if not in the form of state consciousness. I believe that this line of argumentation is correct, but it is equally available to a view which holds higher-order representations to have only this function, e.g. in the form of better self-regulation, not in the form of making experiences more vivid. Phenomenality and higher-order representations thus still remain conceptually disconnected.

12) & 13) Explanatory insufficiency is further highlighted in the arguments by Neander and Byrne, who deal with the existence of higher-order thoughts that are only partly true (Byrne 1997), not veridical (Neander 1998), or not targeted to anything (false positives, Neander 1998; Byrne 1997). In the last case, Rosenthal and Lycan himself admit that there should be something it is like to undergo a mere higher-order state without there being any first-order state at all. Rosenthal takes this to show that, rather than being a vice of HO theories, it is one of their virtues, as it means the ability to deal with misrepresentation. It seems to be the mere existence of higher-order states that brings consciousness about. However, it is difficult to see after these considerations the explanatory force of the transitivity principle as far as first-order states are concerned. Insofar as their relation seems to be this contingent, why is it first-order states that become conscious through their accompaniment by higher-order states? Neander (1998) thinks that this view robs first-order representations of their role in the “division of labor” between the states of handling the qualitiveness aspect while second-order representations are responsible for the consciousness

aspect⁵⁹. Byrne adds that when we have a case of a suitable non-inferential higher-order state without an existing target – and thus consciousness according to HO-theorists – this cannot have any advantage for a higher-order theory over a first-order theory, since both could only tell the story by reference to the process of how one single state came about. Moreover, Byrne argues, on that basis both would be equally unhelpful in explaining why the state would be phenomenally conscious.

14) The other side of this misrepresentational coin is that it seems possible for there to be suitable veridical HO states with proper targets that do not result in consciousness (false negatives). It is possible to imagine a case of *mental censorship*, where certain thoughts do not become conscious because of their hypersensitive character for the subject (cf. Block 2002). These cases of awareness of our mental states would, nevertheless, have exactly the same properties that Rosenthal insists that states that lead to consciousness should have: They are non-inferential, simultaneous and about such states as mental states. Not only do they lack the quality of making their objects conscious, but also these higher-order thoughts even seem to have the opposite effect to their alleged effect in the higher-order theory of consciousness. I think cases like these are perfectly possible, and thus they lend further support to the argument that the HO hypothesis is false.

15) Hardcastle (2004) backs up her argument for conceptual disconnectedness with the remark that there is no empirical support whatsoever for the higher-order view, and it does not explain the differences that she and, e.g., Baars (1988 & 1997) enumerate between conscious and non-conscious states (see section 2.4.). Rosenthal does have a couple of suggestions for the brain mechanisms of HO states and consciousness (Rosenthal 2002 & 2008). One area (medial frontal cortex, Frith & Frith 1999) mentioned as being potentially related to higher-order representing, unfortunately does not correspond to many other neural theories or findings about the neural underpinnings of phenomenal consciousness – or even access consciousness. I think this is a serious flaw in the theory. The other suggestion derives from a result of a study that only concerned binocular rivalry (Lau & Passingham 2006). Connecting higher-order representation with the relevant brain area of the prefrontal cortex (Brodmann area 46) mentioned in that study is rather far-fetched, since the area is usually more connected with attention and working memory. It should also be noted that the higher-order view has been used as an explanatory tool in some neural theories of consciousness (Weiskrantz 1997, Rolls 1999). These, however, have been quite insouciant in their conceptual terminology, and they have often conflated further first-order representations with higher-order representations. Rosenthal himself notes this: “These thoughts are higher order only in respect of having intentional content that is about other psychological states. They are not higher order, as Weiskrantz

⁵⁹ Neander thinks that this problem concerns first-order availability theories just as well, since availability like second-order representations, are down-stream in the sequence of processing.

(1997, p. 72) notes, in being somehow more abstract in content than lower order thoughts." (Rosenthal 2008, 835) Moreover, at least some (Weiskrantz) have refrained from the use of HO theories ever since.

17) Stubenberg (1998) insists that the higher-order model has no resources for differentiating between different forms of non-conscious mentality, such as blindsight or the absent-minded truck-driver, although there is a massive difference between them. If the absent-minded truck driver is phenomenally conscious, as it seems to me most natural to think, this is even worse news for the HO theorists.

18) Animal consciousness is a large topic that looms over higher-order theories. Theorists have two main strategies for coping with this. One is to make higher-order representations all the more primitive (e.g., Gennaro 2004; Rosenthal 1997b), and the other is to deny that animals are phenomenally conscious (Carruthers 1998b, 1989). Both solutions are plagued with a number of obvious problems. It seems counter-intuitive that, e.g., mice have rudimentary higher-order representations that make them conscious creatures. The transitivity principle as an expression of general intuition seems to start to crumble in these cases. It thus seems not to express a necessary condition but rather a typical feature of conscious creatures. I will later argue that this is in fact the case. The other option is to deny that mice are conscious, which seems an equally rash conclusion. According to one line of experimental research, however, metarepresentational capacities seem to extend quite far to the animal kingdom (for a review, see Smith 2003). The question of animal and infant consciousness seems to form a dilemma for the higher-order theorists, and this further corroborates the doubts of the ability of higher-order theories to explain phenomenality. First-order theories have no trouble in explaining animal consciousness.

In sum, the problem that maybe the phenomenality of mental states is not constituted by one's awareness of them is reflected to many specific topics of consciousness studies, and none of the discussions of these topics are advantageous for the higher-order theories of consciousness. The problems related to these specific topics are not easily explained by the higher-order theories. Moreover, the core of the higher-order theories, TP, remains presupposed, not explained.

The other aspect of higher-order theories held as problematic besides explanatory insufficiency and related problems concerns an essential issue related to higher-order theories: self-awareness. It has been maintained that higher-order theories presupposes a form of self-awareness that is neglected or threatens its status as a naturalistic theory, and the higher-order theories must deal with this before they pretend to explain anything (Zahavi & Parnas 1998). Zahavi & Parnas appeal to the long tradition in continental philosophy dealing with the question of self-awareness. Especially, the so-called Heidelberg School, whose philosophical roots lie in German idealism and Kantianism in the Early Romantic Period, for example in the work of Fichte, has dealt with this question. The Heidelberg school has incorporated such diverse influences as phenomenology and analytic philosophy and philosophy of language. All these

influences share an interest in the question of self-awareness. The Heidelberg School can be said to have built their work around one central idea, from which all the main products of it can be understood, and which can be formulated in a relatively simple manner. This central idea is the *critique of the reflection theory of self-awareness*. Higher-order views are examples of the reflection theory.

By the reflection theory the Heidelberg School refers to any attempt to explain self-awareness in terms of an intentional identificational relation by virtue of which we are aware of other objects in general, be it perceptual, reflective, linguistic, conceptual or the like. Any such attempt will encounter the same kinds of problems when trying to explain self-awareness. They will either render themselves viciously circular or else they will lead to an infinite regress. This is because, for the identificational relation to produce *self-awareness*, there must be something that guarantees that the awareness concerns *oneself*, the same subject whose object the awareness is, rather than someone or something else. The seemingly innocent idea that we know ourselves in the same way as any other objects contains this special feature. In the case of the awareness of any other object, the object is different from the knowing subject. In the case of oneself, however, the object is not different, it is the same. And this sameness must somehow be already guaranteed; otherwise the intentional relation would not differ from other kinds of awareness relations, and would not necessarily yield awareness of oneself. It can be claimed that all this is due to the tension between the concepts of *same* and *different*. If intentionality is a relation to something transcendent, something other than oneself, then, when the object happens to be not something different, but the same, this sameness that is special in the case of self-awareness violates this feature of intentionality. Therefore, in the case of self-awareness, the sameness must be guaranteed in some non-intentional manner. This is manifested by circularity or an infinite regress of intentional explanations, depending on the explanatory strategy. Manfred Frank puts the first threat as follows:

"Introspection can only yield knowledge when it relates to something which was already disclosed (and thus already epistemologically accessible); otherwise it would result in a vicious circle: the result would preclude the same conditions it assumed."
(Frank, MS, 4)

In other words, a mental state, which is the object of reflective introspective awareness, must be somehow known to be known as one's own, that is, contain this kind of self-knowledge in order to prevent me from introspecting, say, your mental states by accident. The problem is that in this case, some kind of self-knowledge is already presupposed in the explanation of self-awareness, which renders the explanation viciously circular.

The second threat is entailed if it is presupposed that this pre-required knowledge is intentionally acquired, since it is precisely the intentional relation, which calls for the pre-required knowledge:

"I could not identify any object as myself if I did not already possess knowledge of myself. One who disputed this would find herself in an infinite regress. For her, each

identification of herself as 'I' (that is as herself) would require a preceding self-identification, which would in turn need to be analyzed in exactly the same manner and so on ad infinitum." (*ibid.*)

The threats of infinite regress and circularity loom large over every attempt to explain self-awareness in intentional terms. How are these threats to be avoided? Can self-awareness be explained or is it something primitive that we cannot even start to explain? These questions are also addressed by the representatives of the Heidelberg School. It is not clear, however, whether they can be credited with any satisfactory positive formulation. The Heidelberg School has not done much more than advised the dropping of this form of account in favor of some non-intentional account. The critique of the Heidelberg School is purely conceptual and consists only in *negative* transcendental arguments. It may be seen as a problem that no positive account of this view can be given (Zahavi 1998). There must, however, be *some* kind of self-relation in the reflected act. It is often concluded that the lower-order-reflected act must already be somehow self-aware in a non-reflective or pre-reflective way. Frank speaks about a non-intentional "self-familiarity" or "self-acquaintance" inherent in all mental states. Frank analyzes this pre-reflective self-familiarity further into two phenotypes. First is "experienced phenomenality", which Frank conceives of as the irreducible subjectivity and the self-givenness of the experience itself. The other is the "mineness" of mental states: every mental state is had by some subject. This is best manifested by non-experiential and non-qualitative mental states, but it concerns all mental states in an equal manner. It would be senseless to say that there is pain in this room and ask to whom it belongs. There can be no mental states floating around with no subjects having them.

In my view, it is questionable whether these two phenomena are manifestations of the same thing. In my view, the mineness of mental states does not have to mean a prereflective self-awareness where an experience is aware of itself. Rather, "mineness" means that some kind of self-pole or self-system is present in mental states. Presence of the same self-systems in the structure of reflecting and reflected mental states guarantees the relation for constituting self- rather than other-awareness.

Does the Heidelberg School pose a serious challenge to the higher-order theories? One natural way to dismiss it in this context is to criticize it as a purely conceptual. Perhaps it is our concepts that contain this circularity. Perhaps there is nothing corresponding to them in the real world? How can we know if it is not more than a pseudo-problem that has gained a foothold only in a couple of German armchairs? One learns of oneself in real life just the way one learns about other things, so why make this a problem if not necessary? Could we somehow test this in a methodically rigid manner? What kind of empirical finding would convince us about this? I think that these considerations cannot be directly tested empirically. They can continue to serve as important criteria in theory formation, although they must be susceptible to possible empirical findings that may give us a reason to modify them.

The considerations of the Heidelberg School about the special status of self-awareness are not alien to the analytic tradition. Many of the same points have been made recently by José Luis Bermúdez in what he calls the “paradox of self-consciousness”. The paradox has two aspects, both forms constituted by a circularity akin to the one noted by the Heidelberg School. The first aspect is that any attempt to explain the capacity to think first-person thoughts would be circular. At the explanatory level, circularity is vicious. At the level of development, the circularity is not necessarily vicious anymore: it is not problematic to see developed self-awareness as a result of less rudimentary capacities. However, they also must exhibit the basic phenomena of self-awareness such as indispensability of first-person thoughts and immunity to error through misidentification. I shall describe these in this order.

First-person thoughts (thoughts with sincere uses of the pronoun ‘I’) have one distinctive semantic property: no other thought can replace it unless it somehow includes first-personality. This property was first formally introduced by Hector-Neri Castañeda (1997). Consider the following sentences

Tarja Halonen thinks “I am wearing a red dress”
 Tarja Halonen thinks “Tarja Halonen is wearing a red dress” or
 Tarja Halonen thinks “The 13th president of Finland is wearing a red dress”
 Tarja Halonen thinks that she is wearing a red dress

Now, we cannot move from the first to any of the following without changing the truth conditions. Due to amnesia, Tarja Halonen might not remember that her name is Tarja Halonen and that she was elected by the citizens of Finland to be the first woman president of the nation. Despite amnesia, however, when using the first-person pronoun, she could not make the same mistake of misidentifying herself as when using a third-person description, or her proper name or even third-person pronoun. So, Castañeda proposes, we should at the least mark the third-person pronoun ‘she’ with the reflexive pronoun ‘herself’ or an asterisk ‘*’ (‘she*’) to indicate when this feature of the first-person pronoun is preserved in the third-person pronoun.

Tarja Halonen thinks that she* is wearing a red dress

In that way we can in some sense retain the first-personality in the third-person description, and by the same token, the same truth conditions.

Another distinctive feature of the first-person perspective is that it enjoys the immunity to error through misidentification or the ‘immunity principle’. This is the property of the first-person pronoun that in its genuine use it is not possible for the user to misidentify who the user is. If, with Gareth Evans, we divide third-person judgments into two components, to an identification component and a predication component, in the case of the first-person pronoun the identification component is unnecessary and thus is the question of identification (Evans 1982). If we now compare first-person judgments and third-person judgments made by me, it does not make sense in the first-person

case to ask whether it is really me who thinks. But in the third-person case there is a genuine possibility of the subject misidentifying himself.

As special cases of I-thoughts, Baker's I*-thoughts also enjoy immunity (Baker 2000, Ch. 3). The '*' of the indirect reflexive and the immunity principle are taken to be closely related. It is, in fact, difficult to see what the difference is between them. Instead, they seem to be reflection of the same underlying structure. The father of the immunity principle, Sydney Shoemaker, already suggests something in this vein in his "Self-Reference and Self-Awareness" (Shoemaker 1994, 91). It is sensible to assume that what the asterisk marks in indirect discourse is precisely the immunity that is inherent in the indirect reflexive.

Based on the immunity principle, Sydney Shoemaker has formulated a specific criticism of perceptual theories of self-awareness that comes close to the account of the Heidelberg School (Shoemaker 1994). He thinks that any perceptual model is vulnerable to a misidentification that should not be possible for thoughts whose object is oneself as subject. Shoemaker concludes in a similar manner to Dretske that introspection can only be awareness *that* I am conscious. The proponents of the Heidelberg school hail this analysis. This is suspect in that the kind of conceptual awareness that Shoemaker offers as a solution is equally of the intentional form of self-awareness. Therefore, the threats of regress and circularity should follow for Shoemaker's solution as well.

The challenge posed by these considerations seems important, but it is not clear to me what the transcendental arguments of the Heidelberg School really amount to. The considerations themselves obtain support from linguistic remarks on the semantics of the first-person pronoun. These discoveries are not eventually purely conceptual, since they can be maintained to be observable in all mental phenomena. The immunity principle is a phenomenological one and it is something that is expected by the more conceptual principles. This provides if not criteria (Zahavi 1998), then at least important constraints for a theory of self-awareness.

The most important question is whether these criticisms really affect higher-order theories of consciousness. The criticisms maintain their force only if higher-order awareness of mental states is taken to be the most fundamental form of self-awareness. And higher-order theorists have responded to this criticism precisely from this perspective (Rosenthal 2003). Higher-order theories can admit of self-familiarity for higher-order states and thus remain unaffected. The question of how this can actually be the case, however, is not addressed by the higher-order theorists outside phenomenological tradition, save some intrinsicists (Kriegel 2009; Williford 2006). I shall deal with this question myself in the following chapter.

Baker reminds us that such a move as conceding that HOTs are first-person thoughts, however, makes the naturalization of HOTs problematic (Baker 1998; cf. Schröder 2001). This is so for the familiar reasons that first-person thoughts cannot be substituted for by naturalistic third-person concepts. First-order theories are not susceptible to this counter-argument. Bermúdez

(1998) suggests that some aspects of first-person thoughts are compatible with functionalism, but elsewhere he thinks that Baker's problem for higher-order theories is an acute one (Bermúdez 2001). I agree that although higher-order theories could handle the problems related to self-awareness, it compromises their status as naturalistic theories of consciousness.

All in all, my appraisal is that replete as they are with problems, even if with interrelated ones, higher-order theories do not manage to explain phenomenality sufficiently. For me, this is only a minor setback. As briefly pointed out above, this has not necessarily been their purpose all along. Rather, they purport to explain the aspect of consciousness of there being something it is like to undergo conscious states. This is, in their view, *not* only phenomenality or qualia, but, unsurprisingly the fact that one is at the same time somehow aware of being conscious. Their view is rather underspecified but hints of the following kind are occasionally given:

“But it is no claim of either HOP or HOT per se to have explained anything about qualitative character; they are theories only of the distinction between mental states one is aware of being in and mental states one is not aware of being in. Some other theory must be given of the original qualitative character of a sensory state, such as the yellowy-orangeness of an after-image, the pitch of a heard tone, or the smell of an odor.” (Lycan 2004, 96)

Note also that Armstrong takes “introspective consciousness” to be real consciousness, not perceptual consciousness, which is characterized more like phenomenal consciousness. Even Carruthers seems to waver sometimes whether what he means by phenomenal consciousness is quite the same as it is normally conceived but something in the same vein as other HO theorists think (Carruthers 2000, 127–128). Given this, nearly all the arguments presented against HO theories lose their bite. Despite their lacking the means to account for phenomenal consciousness, they may be able to explain this more self-aware form of consciousness – whatever that means or turns out to be. Unlike the higher-order theorists, I have an explicit suggestion for this, presented in more detail in the next chapter.

4.6 HOP and HOT: a false opposition?

The bottom line is: higher-order representations are wonderful, but they just are not sufficient for explain phenomenality. Taking the criticisms of the higher-order theories collectively and separately, the task is to learn from them and formulate a less problematic version of the higher-order view and try to make some use of it. This will be the task of the next section. Before doing so, I will attempt to solve the debate between the main higher-order theories by discussing, how higher-order representations should be viewed. My contention is that the dispositional higher-order thought view describes a convenient intermediate view that enjoys the benefits of both views. In this way, I will

argue, it is not even necessary to oppose these two views. They just happen to describe the strengths of a less problematic view. The HOP and HOT theories are usually taken to exhaust the alternatives. The fact that both of them have managed to point to several problematic features in the opposite view renders suspect whether either of them constitutes a viable position, the right one perhaps being one that enjoys the strengths of both.

The HOT theory contains many non-thought-like elements. Even though the HOT theory pretends to construe access as thought-like, it has several properties that are contrary to the idea of thinking, as conceded by themselves or pointed out by others (see especially Lycan 2004):

- 1) The access to the first-order state is non-inferential
- 2) The target state is present
- 3) The target state is a potential target of attentional monitoring
- 4) The process is not content-driven
- 5) There is a strong intuition about the existence of an inner-eye

The HOP theory, in turn, pretends to construe access as perception-like internal sense. It contains a few features that run counter this intuition, and is often seen as consisting in quasi-perception, and rather described as self-scanning or monitoring, for the following reasons:

- 1) There is no sense-organ for the internal sense
- 2) There is no sense-quality corresponding to the internal sense
- 3) Some kind of conceptual knowledge is anyway needed to guide the inner sense, and to retain and manipulate the information provided by it

A hybrid position would suffer from most of the problems of the HOP view, and complexity besides, so something more original should be sought. The importance of the HOP–HOT distinction has also been called into question (Stubenberg 1998; Güzeldere 1997; Van Gulick 2000) and more general terms such as monitoring (Güzeldere 1997) and introspection (Stubenberg 1998; cf. Armstrong 1968) have been proposed. These suggestions are insufficient, insofar as they name the problem rather than say anything about the nature of higher-order representations. Again, something more original is called for.

The HOP–HOT distinction can also be questioned on the grounds that there is no principled way of making the perceptual–conceptual distinction. Overall, I have argued that some ways of making the distinction are justified; especially those related to the capacity to discern and recognize abstract higher-order principles in many lower-order tokens. This view then supports the possibility of a conceptual element in the HO theory. One interesting proposal to this effect is the perceptual symbol theory of Laurence Barsalou (1999)⁶⁰. For him, perceptual symbols are symbols that are records of the neural activation of perceptual states. These records are compositional so that perceptual record of one feature in a perceptual image can be activated independently of other

⁶⁰ Barsalou's view strongly resembles the view of Prinz presented in section 4.5, unsurprisingly for the natural reason that Prinz is a student of Barsalou.

components in the picture, and combine with other pictures. Moreover, the visual system includes information-processing that is purely qualitative. Some neurons can fire in the presence of vertices of two lines, irrespective of the lengths of the lines or the angles between them. Then, neural records in the visual systems that code lines and vertices can be activated simultaneously. This activation would represent any triangle, but would not, at the phenomenal level, look as of any given individual triangle for the subject. Barsalou proposes that introspection can work in the same way: “[S]elective attention could focus on emotional feelings, filtering out the specific circumstances leading to emotion, and storing a schematic representation of the experience’s “hot” components” (Barsalou, 1999, 585; cf. Metzinger 2004, section 2.2.1).

My view is that the perceptual symbol theory cannot escape the abstractness of conceptual realm, and it is thus describes genuinely conceptual representations. Barsalou even manages to describe how perceptually derived concepts work. As opposed to Barsalou, I would not say that such activation has to be non-conscious, however. I think that there can be something it is like for the subject during such neural activation. There is a sense of the meaning of “triangle”, even without images of exemplars or inner speech uttering the word “triangle”. The view does, however, show that a principled distinction between perception and cognition is not clear-cut: perceptual systems work like amodal concepts are traditionally thought to do. This considerably alleviates the tension between the higher-order views. The “mental pointing” view of Prinz shares its basic features with this view. However, in both cases I would say that perceptual symbols and attentional mechanisms rely on recognitional capacities that can be counted as concepts. They can be used to retain information and allow it to be used in recognition, attention shifts and further processing in the absence of stimuli. Perceptually derived psychological concepts can be formed in this way. Many may find it counter-intuitive but I think that there is even some mental quality in introspecting, say, one’s experienced emotions by attending to their “‘hot’ components”. I grant the existence of distinct qualitiveness to higher-order states, though I hold that these should be viewed as conceptual.

The perceptual symbol and mental pointing theories are promising alternatives; for me they count as conceptual theories. Furthermore, I think that there is one aspect of higher-order representations, also on this view, that is clearly non-conceptual. The dispositionalist higher-order theory of Carruthers (2000) has been taken to be a version of the HOP theory in that it invokes an analog – picture-like higher-order contents –, which are more akin to perception than thinking (Browne 1999), and which are remarkably similar to those described by Prinz⁶¹. For this reason, Carruthers has himself vacillated between calling his view a HOP or a HOT theory. I think that this is both symptomatic of and lends support to the idea that it is not a version of either a HOP or a HOT theory. It shares important features with the perceptual symbol theory in that it takes working memory to be an important aspect of introspection. However,

⁶¹ Prinz, however, strongly resists his view as invoking recognitional concepts.

what counts as *access to the contents of working memory* is similar to the one described in global availability theories. Such access does not seem thought-like in that

- The access to the first-order state is non-inferential
- The target state is present
- The target state is a potential target of monitoring
- The process is not (necessarily) content-driven
- There is a strong intuition about the existence of an inner-eye

It is also unlike perception in that

- There is no sense-organ for inner sense
- There is no sense-quality corresponding to the internal sense
- Some kind of conceptual knowledge is anyway needed to guide the inner sense, and to retain and manipulate the information provided by it

A higher-order representation view that invokes perceptual symbols as conceptual capacities and global access to working memory as the cognitive access to one's representations is one that is not plagued by the structural problems of traditional higher-order theories. It differs in important aspects from both traditional theories, but carries the advantages of both. These preliminary considerations of a positive account complete my appraisal of the higher-order theories. A more thorough exploration of the potential of HO theories is the remaining task of this study. The higher-order views are summarized in Figure 5 below.

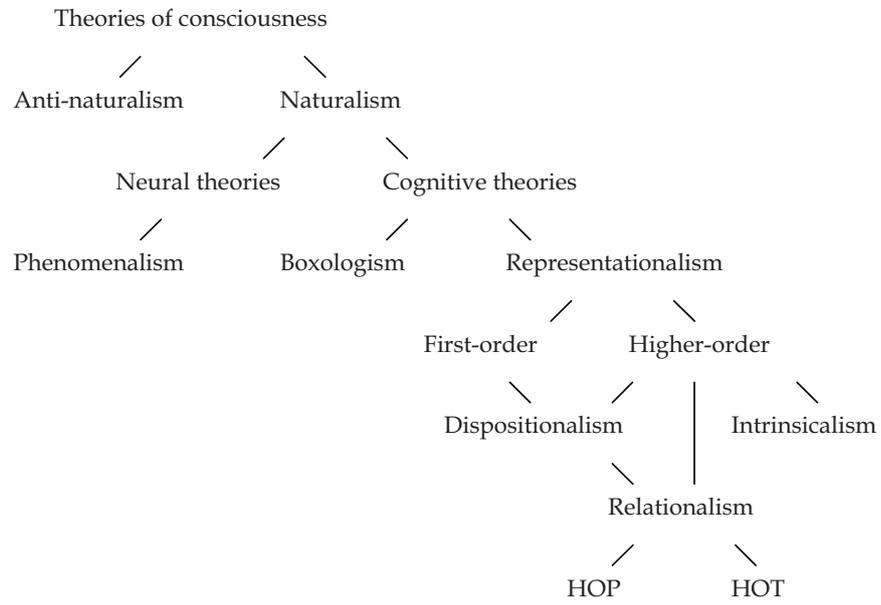


FIGURE 5 The consciousness tree, version 5

5 FORMS AND PROPERTIES OF HIGHER-ORDER CONSCIOUSNESS

In this chapter, I will attempt to show that the debate over the different forms of higher-order theories yields invaluable resources for elucidating the nature of the relation between consciousness and self-awareness and the nature of our awareness of our mental states in general. Most importantly, I will apply the higher-order theories to argue for a distinct form of consciousness, “mentalized” consciousness. Bearing in mind the demands that have been pressed upon theories based on self-awareness as presented in the introduction and the previous chapter, I will proceed as follows. First, I will try to answer the question of what the objects of higher-order representations are. Then I will present the basic psychological properties of higher-order representational abilities: the privileged access to them, the self, the nature of the psychological conceptual capacities needed, and their development. On the basis of this framework, I will try to distinguish clearly between the different forms of higher-order consciousness. Finally, I will deal with a number of possible problems with, and a hitherto unmentioned alternative to my suggestion.

5.1 Objects of higher-order representations

The transparency thesis solves the issue about the objects of introspection for the first-order theorists: there is nothing over and above the properties of objects that we can represent. The issue concerning the object of higher-order representation has been surprisingly neglected among HO-theorists. The case of Rosenthal was shown to be problematic, since he takes mental qualities to be the object of higher-order representing, mental qualities being effects of the qualities of physical objects (the homomorphism theory, Rosenthal 1991). Mental qualities are internalist counterparts of external properties. In the terms used by other internalists, the object of introspection would be a *narrow content*, as sketched earlier by Chalmers as the epistemic content (or mode of

presentation)⁶². I think it is clear that this does not suffice for the present purposes. Although Chalmers ties epistemic content to phenomenology, the contention that epistemic content is shared by intrinsic duplicates entails that it cannot itself be what we introspect. I cannot introspect the inner state of my physical duplicate despite their qualitative similarity. A natural addendum to the narrow content proposal would be the distinction between types and tokens. A type is something abstract, sharable by any number of subjects capable of conforming to the type of mode of presentation. Fregean senses resemble modes of presentations of this kind. They are not, however, sufficient for the purposes of psychological explanations (Fodor 1998), nor as objects of introspection. Tokens of such a type, however, are something that suffice for psychological explanation. Types are qualitatively identical and can be shared, while the tokens of these types are numerically distinct from each other, and they could be objects of representations.

The content of introspection is perhaps not identical with but, nevertheless, can be seen as constrained by the epistemic space. The mode of presentation or epistemic content is clearly a part of one's mental state. One proposition for the link could be that we are dealing here with a contextual intension of the epistemic intention. The context in this case happens to be one's internal state from the first-person perspective. Even if one is an internalist about representational content, higher-order representing cannot consist in observing the narrow contents of first-order representations. That would not amount to implementing the transitivity principle. Narrow contents are often indistinguishable from wide contents in phenomenology. The thesis about the transparency of experience concerns narrow as well as wide contents. For introspection to count as internal, narrow contents should be made to manifest themselves *as narrow* for the subject. HO theories have resources to respond to this. Narrow contents can be conceived of as contents of *my mental states* when proper conceptual capacities are recruited, namely, a set of proper phenomenal or psychological concepts. The object of higher-order representation should have both the worldly content, and also something else that guarantees its being given as a subjectively experienced content.

A phenomenological study, *à la* Smith (1989) not only yields us the token epistemic content or mode of presentation on which one acts but also another important dimension of mental states, namely that of the *modality* of presentation. Whereas a token of a "mode of presentation" is an essentially

⁶² Other theories of narrow content like those of Kriegel and Prinz confuse appearances or modes of presentations with secondary properties, i.e. the subject-dependent properties of objects (Prinz 2000; Kriegel 2008b). These, they claim, qualify as narrow contents. Their account is not narrow enough. Secondary properties, after all, are properties of objects, external things. With this bit of externalism, familiar problems sneak in through the backdoor. Take, for example, a subject first seeing candlelight in normal conditions, and then in another instance hallucinating a qualitatively identical candlelight. She cannot share the secondary properties as the same contents of the mental states in these two instances. When hallucinating, the subject cannot have the same secondary properties of an object as objects of her perception, since she is not perceptually related to any object at all. What follows from Kriegel's and Prinz's view is disjunctivism all over again.

world-directed or self-transcendent element of the mental state, there is the inner component of the modality of presentation that describes *how* one is directed to the content: thinks, sees, doubts etc. This resembles the attitude-proposition distinction, but is more general (Zahavi 2003). Obviously, it also refers to what phenomenologists call the “act” of consciousness. In a case of complex content, such as, “I see a snake”, the snake is the mode of presentation, and seeing is the modality. Mental states are complexes of the two components, or better three components: 1) the “I”, as the subject – that bestows the indexicality on the mental state –, 2) the modality by which the subject is directed toward 3) the mode of presentation.

It is unclear whether they even can occur apart from each other. In affective states, especially in connection with pain, it is sometimes difficult to distinguish the two. Previously, I maintained that there is intentionality in pain as well, even when it is not clear whether there is something else besides the unlocalizable pain itself. In these cases, the mode of presentation is the *world* associated with a property (core theme) specific to pain: an intimate most avoidable palpable harm. The mental state is then at first glance transparent. The modality of presentation in pain is pain sensation.

The interesting question of the location of phenomenal properties gets affected by this. To take color or pain as examples, it becomes instantly clear that phenomenal properties do not belong *exclusively* to either of the components. Color is not simply a property of the modality, it retains its transparency, its tendency to be a property of something in the world, without simply being a property of the object in the world. This is because it is confined to a certain modality of the mental state. Experientially, there are no colors in hearing, feeling or thinking about a snake, but in visual experience as of a snake. A color becomes a visual quality only when conceptualized with its modality. What is called redness in my visual perception is redness of the visual perception when it is conceived as *visual* redness. As long as this is not accomplished, the redness remains the redness of the worldly snake. Even in cases of mere color experiences, they tend to be spatially arranged, or otherwise as belonging to the world.⁶³

In every mental state, the modality is always given with some content. One can not observe mere modalities. Therefore I think that it is the complex mode-modality that is the object of introspection, a complex that has mode and modality as inseparable aspects of one and the same thing. They can only be distinguished by specific means. The modality is not given or cannot be

⁶³ Cf. The view of Chalmers (2004, 28): “When one introspects the content of a belief such as Hesperus is bright, one does so by thinking about Hesperus; one looks right through the mode of presentation. But nevertheless the mode of presentation exists, and one can become introspectively aware of it. The same goes for manners of representation, such as visualness. One looks “through” this manner of representation on introspecting an experience of phenomenal greenness, but the manner is nevertheless there, and one can obviously become introspectively aware of it.” Here, Chalmers defines “manners of representation” as “phenomenologically specified”. Thus they are accessible from the first-person perspective in the same way as what has been called modalities of presentation.

differentiated from the mode of presentation if one cannot make the conceptual distinction. For a creature lacking the concept of sight or seeing, the visual qualities are given as properties of the objects, according to the idea of the transparency of experience. The transparency is circumscribed when the concept of seeing is attained. Then one becomes able to scrutinize the properties of one's mental state in introspection on one's seeing.

One's contact with modality can only be achieved by effort, like in introspection or phenomenological bracketing of the world. One can concentrate and scrutinize the properties of the mental state as different from those of other modalities. Or, one can observe the temporal aspects of the experience as opposed to those of objects. One can then enjoy the phenomenal content of the experience given *as* phenomenal. The *same* perceptual content is now given as mental, by a "cognitive transformation" (Thomasson 2005). Instead of the givenness of a red tomato, the visual phenomenal redness as of a tomato becomes the actual object of your introspection. The source of information is changed accordingly, depending on one's capacities of extraction. One can say one is aware *of* the properties of the given experience, not only *that* one is aware of a tomato. Of course, the first-order state remains directed at the tomato.

Modalities can become passively given when the ability to differentiate between mode and modality and different modalities has become a *routine*. Then, it is possible to develop the intuition that one's mind is immediately given to one. For little children, only things and their properties are given, not their seeing. This may hold beyond childhood, as suggested by the following observation by Lycan: "It takes philosophical sophistication to see that vision really is representation; indeed, some philosophers still dispute the latter thesis" (Lycan 2004). Lycan claims himself to have reached sufficient sophistication to resist the transparency thesis about introspection:

"I can also focus my attention, at will, on further properties of those contents, and properties of the containing state as well. (Thus, I deny the full-blown "transparency" thesis defended by Tye (2002).) I introspect a green patch in my visual field. Let us grant that that is in part to detect external, physical greenness; but that property itself determines nothing about any sensory modality, much less any finer-grained mode of presentation. I also introspect that the greenness is visual rather than something I learned of in some other way. I can also tell (fallibly) by introspection how compelling a visual experience it is, how strongly it convinces me that there is a real green object in front of me." (*ibid.*)

The two components cannot then exist apart from each other. Nonetheless, there is certain privilegedness to the content (mode of presentation). In ordinary perception, it is the content that "makes consciousness intentional, furnishing the act with its directedness" (Zahavi 2004, 22). This is manifest in the fact that we are by default aware of things, the mode of presentation (psychic equivalence). There is no constant *awareness* of the modality, though the modality is there in the experience. Being present in the experience makes it possible for us to become aware of the experience. This, however, requires a

higher-order act with the modality itself as the mode of presentation (cf. Frege 1892). It is only on special occasions that this happens.

In sum, when introspecting one finds a mind with these three elements to it. There is the mode of presentation that is primarily given to one as the object of one's mental state. This time the whole mental state is the primary object and the intentional object of the mental state is only a secondary part of it, the property of directedness of the mental state to something other than itself. But then there is also the subjective part, oneself, the "I", and the way in which the "I" is directed to the mode of presentation. This is the modality of presentation. One can only become aware of this with the appearance-reality distinction in the sense that one can differentiate mind and world, or mode and modality of presentation.

By virtue of there being an "I" or a self-aspect as an element of the mental state, it is guaranteed that introspection yields *self-knowledge*, where something different, the mode of presentation is also given. The modality of presentation is actually something that can be said to knit together the self- and other-aspects, and specifies their relation. Interestingly, when moving from a first-order presentation to a second-order presentation (introspection), the other-aspect is phenomenologically downplayed in a sense. In introspection, the same "I" or self-aspect is extended to cover both states, the first-order and the higher-order states. The mode and the modality of presentation change, but the previous modality becomes the mode of presentation. The modality is more characterized by self-aspect. Hence, there is less of the other-aspect in the mode of presentation, and the result is self-awareness. In my view, this is by virtue of the "I", the background systems, and its property of remaining (relatively) constant as compared to the first-order representation across different experiences. Thus, guaranteed self-knowledge, things like immunity to error through misidentification can be expected just as in other kinds of first-person thoughts. The "fracture" into self- and other-aspects, and the resultant complexity insisted on by Zahavi (1998) is thus made clear and unified by these components, and they can be embedded in the dispositional higher-order model. In that model, the succession of modes of presentations is short term memory, as described in the global availability theory.

The object of higher-order representation should be seen as the occurrent token complex of mode and modality of presentation. The question remains whether all kinds of mental occurrences can be subsumed under this view of introspection. It has been proposed that introspecting is radically different in the case of thinking and perception (Carruthers 2004), and perhaps this view applies only to some mental phenomena but not to others. I believe that it holds for all conscious mental occurrences, in addition to the previously mentioned perceptual and conceptual states, and even to affective and conative states.

If this view about the objects of higher-order representation is adequate, then introspecting emotions should be analyzable in the same manner. Emotions can be seen as representational, that is, they have a representational content (mode of presentation) and modalities specific to them. In terms of evaluative intentionalism, in emotions the representational content is

something with value⁶⁴. Modalities can be distinguished on many levels, such as those of specific emotions, fear, love and so on, or only of emotion more generally – or even only the modality of a phenomenal experience, as suggested by Smith (Smith 1989). I rejected earlier the view in which no distinction is made between, e.g., being in pain and being aware of being in pain, or that affective states do not have intentional objects. I think this simply is a misdescription of pain and affective states. It becomes obvious once pain and affective states are considered in terms of their modalities. The modality of pain belongs to sensory and affective modalities, while the modality of awareness of pain is not sensory or affective but rather conceptual thinking or believing. As pointed out earlier, instances of seemingly objectless pains do not lose their self-transcendent aspect of mentality, and they have at the minimum the whole world as the mode of presentation.

Conative states such as desire and volition can also be analyzed this way, even though they constitute an important exception in one sense. There are influential thinkers who have maintained that some aspects of our self-awareness, for example that of volition, can only be described *nachträglich*⁶⁵. In these views, volition is non-representational and elusive, so it cannot be directly metarepresented and observed. Moreover, it has been maintained by some who share the spirit of the representationalist theory of mind vindicated here that desire is elusive in that only its effects can be perceived let alone that it can be directly introspected (Schroeder 2004). Conative states cannot be reflected upon as occurring.

However, in the context of representationalism, a conviction exists that both desire (Schroeder 2004) and volition are representational (Prinz 2002; Mandik 1999). In the inspirational suggestion of Mandik, it is the effects, not the causes that are represented. Somewhat paradoxically, then, the representation exists before what is actually represented. This seems to contradict the causal analyses of representation. Mandik differentiates effect-based representationalism from causal representationalism and calls the former procedural information in contrast to the more traditional causal information represented by the latter. Conceiving of volition in this way suits well higher-order representational views of the introspection of volition. A representational analysis of volition allows a two-dimensional version in that one kind of content is internal to the system that perhaps only requires causal interactions within the system. The other content is determined by external factors, in the case of volition the action performed by the agent. Interestingly, the question of truth and falsity is difficult to locate: what is it that determines the truth of volitional representations? Is it correspondence with muscular movements or

⁶⁴ See (Seager 2002) for a good representational analysis of emotion and emotional introspection. It differs from mine only in that it sees introspection in accord with first-order representationalism as awareness that, not awareness of.

⁶⁵ French philosopher Paul Ricoeur has written a couple of books based on this idea arguing that transcendental phenomenology cannot capture volition and action, but that an indirect moment of interpretation (hermeneutics) is needed (Ricoeur 1960a & 1960b).

the outcome of the action? Shaun Gallagher's reinterpretation of the results of Benjamin Libet's classical experiments might be helpful here (Gallagher 2005, ch. 10). Conscious intentions and volition operate at the level of worldly contents such as getting a cup of coffee. The movements that are required to achieve this goal, in turn, are controlled automatically by action scripts that constitute our body schema, i.e. bodily representations that guide action but which cannot become conscious or objects of representation. Both may allow a distinct representational analysis. Only the first, however, is accessible for introspective metarepresentation. The body schema is interesting with regard to the question whether it is phenomenal itself or not. If so, the analysis of phenomenal consciousness defined as the potential contents of access consciousness needs to be revised.

5.2 Psychological issues in higher-order representations

Specifying the contents of higher-order representations is crucial, but I have not said anything yet of the psychological underpinnings and neural mechanisms of higher-order representations. Although only an initial sketch is possible here, both are required for a comprehensive evaluation of the potential of higher-order theories of consciousness.

At the cognitive level, at least three basic psychological components have to be accounted for, as stated in the introduction: psychological concepts, which enable the attentional observation of the occurrent states of the mind, working memory to enable attention and access, and the self-systems that guarantee the indexicality of the object mental states of higher-order representation.

All these can be characterized in the context of the global availability view (Baars 1988, 1997), which is a *first-order* availability theory, but on a *higher-order* version of it, one of the non-conscious systems is dedicated to higher-order representational capacities (Carruthers 2000). My conviction is that a first-order view should suffice for consciousness, given the problems of the higher-order views. A first-order view can, however, incorporate higher-order reflective capacities irrespective of consciousness. These capacities could be just one among the non-conscious systems that determine the contents of consciousness. By these capacities it is possible to explain introspection without invoking the implausible displaced perception view of the first-order theorists. The functioning of such a system can of course become so automatic, that it can also explain the *intuition* of higher-order awareness of consciousness as a permanent and essential feature of consciousness.

Availability and psychological concepts are familiar enough by now from previous sections (2.4 and 4.6, respectively). Access, in turn, is simply access to working memory. This access is privileged, and has the characteristics that correspond to an inner sense view. It should be noted that in the case of higher-order representations, the temporal dimension is more important than in other

cases. One good way to understand this is through the Husserlian analysis of time-consciousness (Husserl 1966). In Husserl, the retentive aspect includes not only retention of the object but also retention of the mental act. One's access to mental states is access to previous states that can be retained and observed in the working memory. Shaun Gallagher's characterization of the retentive aspect of Husserlian time-consciousness gives conceptual unity to this proposal (see n. 56).

Describing such self-systems deserves, however, a more thorough examination, since they hold the key to understanding the indexicality and "mineness" of mental states. However, it has become an increasingly popular view that, we have no better evidence that such things as selves exist than do elves (Dennett 1991; Blackmore 1999; Metzinger 2003). Postulating selves means for Dennett postulating a "center" in the brain where all conscious contents are gathered, which further requires postulating a self or *homunculus*, to which the contents are represented. This Dennett calls the "Cartesian theatre" view. What evidence we have of selves is only evidence of cultural and representational constructions. I think this is partly right. I do not claim that selves are things. They are representational structures of experiences, not some kind of substances. It is true, however, that these arguments have wholly exorcized the question of self from the philosophy of mind for a considerable period of time, and I agree with Baars and Damasio that these arguments have led to an unnecessary hypersensitivity to anything related to the concept of self.

As a general strategy for an account of self-structures, I will adopt the view of Baars discussed in section 2.4, that of the self as *the deepest context* of consciousness. Personally, I have no problem calling the various self-systems and self-aspects of the experience the self, without invoking any distinct metaphysical substance or entity. There is a whole person who has self-structures that are shared by many of the types of mental states that the person enjoys. The various self-structures forming the deepest context of consciousness are the relatively stable or invariant background structures in contrast to the richly variant contents of consciousness. Baars already enumerated personal goals and autobiographic memories as important self-structures. One important and developmentally perhaps the earliest self-structure is the body and its representations in the mind, as mentioned briefly by Baars (Baars 1997, 146–147). Many of the self-systems have a bodily origin. Put in Gallagher's words, "[t]he human body, and the way it structures human experience, also shapes the human experience of self, and perhaps the very possibility of developing a sense of self. If the self is anything more than this, it is nonetheless and first of all this, an embodied self" (Gallagher 2005, 3). The earliest forms of self include also the "constitutional self" described by, e.g., Fonagy et al. (2002, 11), which refers to the biologically determined and bodily propensities to react to things strongly or weakly by one's autonomous nervous system, or temperament in a word.

At the neurological level, numerous self-systems have been identified. The most important theory in the present context is that of Damasio (1994 & 1999), who has a theory that describes the mechanisms for a self that specifically

emphasizes the here important features of self-systems being invariant, contrasted with and boosting the external contents of consciousness. The theory shares the primacy of the bodily self by discerning two kinds of self. The primary self is the “core self” which consists in representations of the internal milieu of the body, the input from the proprioceptive and interoceptive senses, the first dealing with the perception of the state of the skeletal muscles, the overall position of the body, and the second involving sensory input from involuntary body parts, the visceral systems, cardiac tone and so on. This sensory information is processed at various levels in the central nervous system, from the pons up to the somatosensory cortices (proprioception) and the insular cortex (interoception). The areas dedicated to monitor the body and related to the body include parts of the brain stem, like superior colliculus, which matches movements of eyes and head to the visual field. What is noteworthy is that the areas constituting the bodily self are also ones identified by Damasio and others as the neural mechanisms participating in production of emotions. The bodily self is also an emotional self. This is compatible with the trait theories in personality psychology invoking emotions or dispositions to them (temperament) as central characters of personality. The constitutional self, the biologically determined features contributing to emotion dispositions and temperament characteristics of the person, fits this view of the self rather nicely. These stable characteristics in addition to bodily representations constitute the primary forms of self on the developmental and experiential levels. This core self is short-lived and is rebuilt in every episode of perception. The narrative and teleological aspects of the extended self build on the core self, with the aid of the working and episodic memories. In perception, a map of the internal milieu related to the core self is contrasted with the sensory images supplied by the outer senses. Damasio speculates that this division of consciousness into other and self accounts for the directedness of the mind to outer contents (Damasio 1999).

The Gibsonian ecological theory of perception (see n. 22) extended the term proprioception to include the external senses. All of them also serve a proprioceptive function. The boundaries of the visual field provide information about the body, and temporally of the movement of the perceiver through the environment. Also “affordances” as indexical properties of the environment provide information about the perceiver itself. The theory of Damasio does not take this completely into account, though he includes vestibular systems and the superior colliculus, related to eye-movements and stability of the visual field, among the self-systems. The theory, however, is easily complemented accordingly. All this is compatible with the view that there is an invariant self-aspect in all perception. It would be interesting to explore the potential of this thesis to generalize across all mentality. Considerations of the impossibility of a principled distinction between perception and cognition would open an interesting avenue for such an exploration. Phenomenologists’ description of an ‘I’ accompanying every thought is of course another aspect of this issue.

Critics of Damasio’s theory (Panksepp 2003; Craig 2005) have stressed the importance of the sensori-motor systems and motivational and emotional

executive systems as components of the self systems. This criticism is justified, and Damasio does in fact do this, as he stresses the cingulated cortex as an important self-system. The anterior part of the cingulated cortex is related to emotional motivation, and massive lesions in this area manifest in akinetic mutism, a literal cessation of self-initiated action or even of responding to the environment. Only very intensely agitated, a patient suffering from this may respond briefly by, e.g., pronouncing her name but rapidly returning into her extremely passive state. In addition to the motivational ones, the systems of the brain in the frontal cortex related to intentional and planned action are important in the production of the goal-related aspects of the self. This is also nicely described by Damasio (1994) by reference to the notorious case of Phineas Gage whose personality was dramatically altered after the protrusion of an iron bar through his frontal lobes, especially brain areas related to intentional action and personality traits.

The motor systems also contribute to the sense of self. The motor systems of the basal ganglia are speculated by Panksepp (1998) to produce a feeling of ability or the sense of possible movements performed in a situation. It is known that the basal ganglia are an inhibitory system by default that works by allowing some motor signals to get through to body parts. All the time, many motor urges elicited by sensory perceptions and internal motivational systems are relayed forward from the motor cortex but inhibited by the basal ganglia⁶⁶.

All in all, the distinct self-systems do not constitute an independent entity or substance residing in the brain. Rather, they form a contrast to outer perception, the relatively invariant elements of the mind. Together, they form a fracture in the mind, a division into the explicit and marginal intentional contents of the mind on the one hand, and the implicit self-structures on the other. These self-structures can become intentional reflected contents of consciousness to the detriment of other self-systems. In Gallagher's terms, there are always body schematic structures implicitly supporting the explicit body-image, i.e. body representations that are reflected upon (Gallagher 1995).

I suggest that all mental states derive their first-person form from self-systems, especially the body. Zahavi explains these by his notion of pre-reflective self-awareness (Zahavi 1998), which he analyzed as awareness of mental states. In my view, this is unnecessary, and Zahavi himself concedes the problemacy of any talk of awareness of mental states in this connection (1998, 76). Once mineness is understood by the presence of some implicit self-structures in various mental states, mineness is simply a consequence of the *indexicality* of mental states. Suppose that the relatively invariant structures of an experience, in contrast to the manifold contents of the outward directed consciousness, constitute the self. The presence of such structures would suffice to guarantee the belonging to the same self the act of reflection and the act reflected upon, since it would be the persistent underlying structure from the

⁶⁶ If this is right, the basal ganglia are the mechanism behind the primary self-system of "I can" described by the phenomenologists (Merleau-Ponty 1945; Ricoeur 1950). The "I can" (*je peux/je veux*) is a phenomenological shorthand for the bodily acting subject and a substitute for the non-material Cartesian thinking subject.

one to the other. Zahavi interprets indexicality to be the pre-reflective self-awareness of an *experience*: if one is aware of objects in relation to oneself, one grasps one's unique perspective on it (Zahavi 1998, 24–25). I think that this interpretation is too stringent. There are cases of indexicality that require no awareness of experience. When a rabbit sees a fox, a fair description of the phenomenological content of its perception would go as follows: "there's a dangerous fox, and I must flee from it immediately". To invoke the concept of seeing in the content, as a reflexive content theorist would have it, is an unnecessary complication. If we accept Gallagher's description of our first capacities of social interaction being non-mentalistic skills, and no doubt we and Zahavi himself should, then our first social interactions take place without psychological or mentalistic concepts. In the present case, it can be said that the information of it being an instance of seeing is present in the perception, but the rabbit lacks the capacity to make use of it. The indexicality of the perception would, nevertheless, guarantee the "mineness" of the perception if the rabbit were suddenly endowed with such a high-level capacity of self-reflection.

Indexicality also guarantees the indispensability of the first-person pronoun. If the meaning of the first-person pronoun is attached to the implicit invariant self-structures from the first-person perspective, it becomes understandable why it differs in meaning from whatever other way of denoting the subject. Names, other pronouns and descriptions are not indexical in the same way. In experiences, "I" is "I as subject", not as object, so immunity to error through misidentification with respect to the first-person pronoun naturally falls within it.

Psychological concepts were briefly introduced in section 4.6, but a closer consideration of their possible neural mechanisms is still called for. Nicholas Humphrey, a brain scientist who was among the first to document the phenomenon of blindsight⁶⁷ and who subsequently went to Africa to become an ethologist, was puzzled by the fact that there was no known function for half of the gorilla brain. Observing a population of gorillas, he started to pay attention to the fact that the gorilla's world is half social – just like the human world. He got the idea that the mysterious half of the brain must be for coping in the social world. Humphrey conjectured that through evolution man and other great apes have become so dexterous in sorting out the behavior and intentions of one another, and in consequence he started to call them "natural psychologists" (Humphrey 1986).

Mere conjecture apart, one important piece of evidence in support for distinctness of sensory and phenomenal concepts was obtained by the Danish researchers Morten Overgaard & Thomas Sørensen (2004). They observed that when one concentrates on a feature of the visual environment, ERPs (evoked response potentials in EEG, in response to sensory stimulation) the latencies of the brain responses were shorter than responses to identical stimulation when one concentrates on the same visual feature, say, redness but now as a feature of one's experience rather than of the world. They took this to be evidence that

⁶⁷ See (Weiskrantz 1986).

the difference between introspection and perception is not merely a conceptual difference but a real difference. By a conceptual difference, however, one could mean conceptual abilities, and to take the difference to derive from the use of phenomenal rather than sensory concepts. This is, in fact, what Dretske himself took the case to be. Nevertheless, the experiment provides constraints on how we should think of phenomenal concepts not only on neural level but also on the cognitive level: they seem to require a longer time than mere sensory processing.

Recent research has not confirmed Humphrey's conjecture that half of the brain is required for psychological capacities, but a fair number of brain areas have been associated with higher-order representational systems. There seem to be three brain areas that have systematically been found to be activated in connection with mentalizing function: Ahmad Abu-Akel differentiates between the representational posterior regions (superior temporal sulcus) and some of the applicative or executive regions of the frontal brain (Abu-Akel 2003; Frith & Frith 1999). The executive areas reside in the medial frontal cortex and comprise the orbitofrontal cortex, an area related to affect-regulation and the anterior cingulate cortex, and thus associated with the motivational component of affective states. To the posterior regions belong not only the above mentioned areas but also the amygdaloid circuit. This circuit is essential for some emotional responses as well as the perception of emotion. Other areas related to the perception of emotion are the insular cortices (Freedman & Gallese 2007).

The above mentioned areas are related to higher-order representations (or ToM or mentalization) in general. However, a recent study managed to associate one specific area with reflexive access to one's own states (Vogeley et al. 2001), the right operculum, an area of the cortex in the junction between the parietal and temporal cortices. This area has been related to the representation of self also elsewhere, especially self in movement through space (Tsakiris et al. 2007; Bonda et al. 1995). However, the task was to understand psychological concepts as related to the first person pronoun. This does not quite correspond to the concept of introspection as differentiated earlier, as focal awareness of one's current mental states as mental states, but it is possible that this area might serve as a self-system for mental states and the access to them in general as suggested by the experimental data referred to shortly.

These experiments are only the beginning of a new field of research, but they form the primary sketch for the neural mechanisms in the multi-level model of higher-order representations. New more refined experiments are being run where different aspects and modes of higher-order representation are distinguished. The significance of this research here is also that it corroborates the fact that areas related to higher-order representations seem to differ from those related to first-order phenomenal consciousness. They also tend to support the idea that no internal sense exists but higher-order representation rather requires the conceptual capacities.

5.3 Mentalized consciousness and other forms of higher-order consciousness

Psychological self-awareness is traditionally conceived of as introspection. Introspection alone hardly covers the abilities of higher-order representation. For a more comprehensive analysis, I distinguish between two basic sorts of psychological self-awareness using the concepts and analyses given above. The first mode is that of reflexive consciousness. More specifically, by reflexive consciousness I mean all kinds of occurrent consciousness that have one's own mental states as their content. There are two main sorts of reflective consciousness. The first is that which is directed to one's co-occurrent mental states which are retained in the working memory. This relatively retrospective consciousness of one's own mental states is what I take introspection to be. The second sort of reflexive consciousness, that takes the form of indirect inference, instances of recollections from the long-term memory of previous instances of introspection, is a distinct, more retrospective reflexive consciousness.

Indirect and retrospective reflexive consciousness is not shrouded by any veil of mystery, since it does not necessarily involve any special access to one's states. Introspection, however, still is. As is well known, introspection was an important method in psychology from the 19th century to the beginning of the 20th. A number of the greatest early psychologists, such as Franz Brentano, Wilhelm Wundt, William James, and Edward Titchener, are often called "introspectionists"⁶⁸. The introspectionists based their psychology on the mind's ability to turn on and observe itself in operation. The method of introspection was subsequently discredited by behaviourism. Nevertheless, the introspectionists had the opportunity to produce a considerable amount of experimental data by introspecting and also lots of prescriptions of how scientifically approvable introspection should be practiced. However, it is surprising that the introspectionists never really managed to spell out what constitutes the psychological nature of introspection itself. It was largely taken for granted, as the famous phrase of William James suggests: "The word introspection need hardly be defined – it means, of course, the looking into our own minds and reporting what we there discover" (James 1890). Edward Titchener seems to be the only exception to address the question of what introspection is from a psychological perspective (see Titchener 1912a & 1912b). However, after many years of research, Pierre Vermersch was forced to describe even Titchener's contribution in the following manner:

"In the final analysis the harvest proves disappointing, a fact whose significance he [Titchener] tries to reduce by pleading the novelty of this step and the need of improvement but which leaves entirely unanswered the question concerning the nature of introspective practise." (Vermersch 1999)

⁶⁸ For a good historical review that directly pertains to higher-order theories see (Radovic 2005).

Despite this fundamental failure, Vermersch continues:

“In all the literature bearing on introspection I have not been able to find a position more clearly and lucidly expressed than that offered by Titchener”. (*ibid.*)

Vermersch insists that what is needed is “an introspection of the introspection itself”. I agree but also think that a more comprehensive, multidisciplinary study of introspection is what we really need. Higher-order theories offer an important perspective on the nature of introspection, clarifying issues about access, the conceptual underpinnings and the self-systems required for it. A case can be made by now for the need of language to describe introspection. It was concluded previously that higher-order representation require conceptual thoughts using psychological concepts. Mental states that are the object of introspection are states that have to be retained and rehearsed in the working memory. These mental states can include perception or perception-like representations like mental imagery, or more abstract representations such as concepts. Sequential thoughts can be composed of either. The contents of these states are accompanied with information about their “personal-level vehicles” that, following Smith (1989), I called modalities of presentation.

Psychological self-awareness is not exhausted by this sort of reflexive consciousness. I believe that there is an important phenomenon that has not been taken into account properly in the current debate over the problem of consciousness. This is what I like to call “mentalized consciousness”. This concept of consciousness is based on the concept of mentalization as described shortly in detail. Mentalization or reflective functioning is based on what was earlier called theory of mind, folk psychology, psychological concepts etc. Yet, it is more than this or what philosophers mean by higher-order representations. The recent coining of it that links it to attachment theory is the relevant one here (Fonagy et al., 2002). In this view, the ability to mentalize is less than an abstract way of representing one’s psychological states. It is a context-bound ability to use psychological information to regulate one’s behavior and internal states. It is central in social coordination with other persons and to one’s capacity to relieve anxiety by fantasizing things in a psychological and social dimension.

Different forms of higher-order consciousness are best distinguished with respect to the developmental dimension of higher-order representation, and it is good to be clear about the temporal order of development. There are many suggestions for how one comes to have higher-order consciousness. These include innatism (Rosenthal 2004, 23; Tye 2003, 18; Loar 1997, 602)⁶⁹, and genetically determined theory of mind module (Baron-Cohen 1995). Other suggestions are misperception (Rosenthal 2006), practical social interactions (Gallagher 2005), acquisition of language (Bermúdez 2003) and the interaction

⁶⁹ Husserl’s theory of internal time consciousness can be interpreted such that longitudinal intentionality is an innate feature of consciousness. It is, however, not clear at which developmental stages it starts to operate. If it applies to working memory as Gallagher (2005) suggests, it will be in place almost problematically late, depending on what kind of developmental theory one is prepared to follow.

between an infant and its caregiver (Fonagy 2002). The most important suggestion for distinguishing different kinds of higher-order representing is the last mentioned, but I shall be drawing on multiple sources.

At the beginning of the life of a human being, I think that no abilities of higher-order representation exist. Newborns probably undergo phenomenal experiences, not necessarily like the structured ones of adults. The capacities of being aware of one's mind are at most latent at birth. Therefore, I think that radical nativism is not true. The child's mind is in the very beginning of life characterized by the psychological principle of "psychic equivalence", the qualitative if not numerical identification of inner reality with external reality. The principle can be put as follows:

(PE): Originally, only *world* is given to the subject⁷⁰.

The principle states that there is no distinction between mind and world, or inner and outer realities in consciousness at the outset. Everything in experience bears the full weight of worldly reality. There is no question of conscious experience being wrong – or better, no question of conscious experience. Furthermore, one cannot tell primordially whether the content of the experience arises from within or without, insofar as no such distinction can be made. What is often taken to arise from within, be it pain, hunger or any other affective state, is experienced as a property of the world among external objects. Even though the internal-external distinction is not made, there is a certain bias towards the external world. Affective states extend from the self to the world, coloring the experience of the world according to the given affective state. The *world* is a negative place if one is dissatisfied, positive if satisfied.

Needless to say, the principle of psychic equivalence is compatible with the first-order representational theories of consciousness and the phenomenological principle of transparency of experience, being more its psychological and developmental counterpart. The principle of psychic equivalence states that the content of mind and world is equivalent at the experiential and psychological levels. There are *no appearances*, only worldly things given as they are *an sich*. It is quite plausible to maintain that this is a primitive feature of consciousness, linked to the transparency or diaphanousness of consciousness. Psychic equivalence is thus the *default* mode of consciousness. It is only later that appearances are given to consciousness. This, however, requires a way of revealing them, the psychological capacity to distinguish objects from their appearances. One standard way to explain this capacity has been the possession of phenomenal concepts. Without any corresponding account of it, explaining the givenness of *consciousness* becomes intractable.

⁷⁰ The formulation is mine. Fonagy et al. (2002) introduce the concept of psychic equivalence in the context of developmental pathology. They do not elaborate the concept so much in terms of the philosophy of mind as from a clinical and developmental viewpoint. Nevertheless, I believe that my version shares the essential features of the original one.

Aydede and Güzeldere (2005) give an interesting account of the development of phenomenal concepts, concepts concerning phenomenal properties of experience, from their sensory equivalents, concepts concerning properties of worldly objects. The phenomenal concept RED_p is acquired on the basis of the sensory concept RED_s. Pain is for the authors an instance of a concept where the phenomenal non-sensory dimension is the most vividly present because of the evaluative component. Thus pain most easily makes the transition from a sensory to a phenomenal concept, also among the concepts that are used first as sensory concepts, like from RED_s to RED_p. I think there is much truth in this and that pain certainly contributes to learning of the A-R distinction. Still, I think that this evaluative component cannot guarantee the concept's purely phenomenal nature or that it contain the A-R distinction in itself. I concur with the view that pain as well as other affective states is first experienced in the mode of psychic equivalence. A pain in the foot concerns for the infant first of all *something stinging in the foot*, even if such pains that do not unequivocally locate themselves in the body or somewhere in the world were to exist. In such cases, the infant does not consider the pain to be its mental property at all; instead, the whole world is present to the infant as painful "tinted with agony". The same holds for moods and other affective states: in depression the world is seen as overwhelming, while when in a cheerful mood the world is seen as an agreeable place and so on. Thus, the worldly bias of PE concerns affective states as well.

It certainly seems true that people are primed for social interaction at birth (Gallagher 2005; Meltzoff & Moore 1995). First, interaction is presumably based on imitation and the timing of responses by caregivers to the child's own responses. This interaction between caregiver and child has been well documented and analyzed in attachment theories (Bowlby 1969). Fonagy et al (2002) use this data in their theory about the development of the child from the mode of psychic equivalence to possession of the A-R distinction, or in their terms via "the pretend mode" to what they call the mode of "representational mind"⁷¹. They tell a psychodynamic story of how in addition to biological factors, a child's early social relationships have a determining role in the child's increasing abilities to understand the mentality of other people and herself. The caregiver's ability to mirror the infant's inner states in a correct manner contributes to the infant's developing ability to distinguish its inner states from reality and other people's inner states, and thus to break the spell of psychic equivalence. The infant internalizes the responses of the caregiver, and these internalizations constitute the first means for the infant to regulate her inner states. If the responses of the caregiver are successful, the models of action and emotional reaction are stored as implicit models of self-regulation and interaction with other people. Skills at this level cannot necessarily be remembered, since they develop at a stage that precedes the development of explicit memory, and for this reason cannot be repressed (Mancia 2007).

⁷¹ The term representational mind is borrowed from Perner (1991), and might not be a entirely happy one since it refers to the *metarepresentational* capacities of the infant.

However, it is maintained that these skills can be accessed and thematized in therapy, e.g., in transference in the therapeutic relation and in interpretation of dreams (*ibid.*).

Later on, the capacity for *pretence* gives the infant the ability to gain better control over its inner states by externalizing them in play. In pretend play, the child is aware of the contents of play as unreal and internal to the frames made up by the child's imagination. In this phase, there are only the modes of full reality (psychic equivalence) and full unreality (pretence). At least at this point, a primitive form of the A-R distinction must be in place. In pretence, things are taken to be unreal, not perhaps representations or appearances of existing worldly objects despite their obvious relation to the real world. Only later do these two modes merge into one when finally the child learns that inner states can be partly true and partly untrue. When integrating the modes of pretence and psychic equivalence the child effectively acquires the mode of "representational mind", the understanding that other people's mental states are representations that do not bear the whole weight of reality. Here, the child learns the A-R distinction head-on, as applied in standard A-R distinction and theory of mind tests.

The mental functioning of a normally developed person equipped with normal functioning of "representational mind" is, in the terms used by Fonagy et al., "mentalized", which means that her mental functioning contains automatized self-regulative capacities. A person without such capacities would have no way to control her behavior in the same way as a normal person and so she would remain at the level of psychic equivalence. A person may have local deficits in her self-regulative capacities, "psychotic islands" that are manifested in retreats to psychic equivalence in certain contexts in which the self-regulative capacities have not had the chance to develop normally. This view of mind, it seems to me, has the resources to better explain why there should be awareness of itself in consciousness, and why a normal mentalized functioning of consciousness would develop an intuition of the givenness of consciousness. The infant starts to match its internal states to its caregiver's responses and they become their first internalization and regulative resource. The child learns to anticipate the caregiver's responses, and also internally go through them in the absence of the caregiver, thereby relieving anxiety.

The child also discerns patterns in its caregiver's actions that it can attune its own actions to, for example constrict the muscles of the neck when lifted. Patterns of behavior become subsequently more complex and the child starts to discern intentions in other people's behavior. There is evidence that this ability is reached already before the second year, approximately at the age of nine months (Meltzoff 1999). Also, the child engages for the first time in active communication with its caregivers in the form of joint attention. All this while, it could, however, be maintained that the child remains in the mode of psychic equivalence, insofar as the contents of the affective states may be seen as external.

The next important achievement in the development of higher-order functions is the ability to pretend, the "pretend mode" (Fonagy et al, 2002). At

this stage, there must be some rudimentary ability to distinguish between appearance and reality, since pretend contents are conceived as not being real. However, this conceptual ability is not ready for the detached use required by false belief and A-R tasks. This is well in line with the idea that concepts are not abstract in the sense that the context of use would not affect them (Barsalou 1987 & 1993). Even the detached use obtained between three and four years of age does not amount to what is required by introspection, in my view. The mode of representational mind, ability to see other people's internal states as true or false is also strongly bound to the child's coping in her social environment.

This account is based on the current state of the art of developmental research and is likely to be revised and refined in the following years to come. Therefore, it is also more accurately described and more plausible than the abstract awareness of one's mental states described by the classic higher-order theories. The ability to introspect one's current states is something far beyond that and possible only after the training of use in this sense. It is difficult to say at what age humans in average attain the ability of introspection. My conjecture is that it is not even the privilege of most people. This is perhaps why the early introspectionists would only trust their subject's reports after intensive training periods. In this sense, the ability to introspect one's internal states does not differ from other context-bound conceptual abilities. Theorizing with psychological concepts and their use in strict, well-conducted introspective experiments or in phenomenological description is an extraordinary achievement that requires a long personal history. The concept of mentalization, therefore, does not correspond to this explicit, thematic reflective function, but to the often slowly developing, implicit, practical way of using psychological information in every-day life.

This sort of self-awareness Fonagy et al. described by the concept mentalization on the basis of attachment theories, is, in my view, one that is also essential when making sense of the interrelation between consciousness and self-awareness. This can be seen as a distinct form of consciousness in its own right, one that would be most faithful to the etymology of the term consciousness. I call this kind of consciousness "mentalized consciousness", and distinguish it from "reflexive" consciousness in that reflexive consciousness requires active thematized reflection on one's present or past mental states, but mentalized consciousness requires only the possibility to reflect on one's mental states and implicit guidance by one's psychological abilities.

Having practical self-regulatory background knowledge about one's own mind means control over one's own mind and actions, and the ability to report readily on them. It is plausible that this sort of consciousness requires the transitivity principle and is close to the conception of consciousness that a layperson would have. It is also different from introspection and other forms of reflexive consciousness in that psychological information does not form the primary content of consciousness. Rather, it conforms to a dispositional higher-order thought theory or even an intrinsicist view in that it includes the possibility of becoming explicitly aware of one's mental states. It differs from

existing dispositional and intrinsicalist versions higher-order thought theory in that it is more practice-bound. It contains the idea that one's behavior is actively regulated by information about the social and psychological aspects of one's current situation although these are not the occurrent content in one's mind. Rather, reflective functioning is automatized to the extent that no reflection is needed in order to act in conformity with the psychological demands of the situation.

I suggest that this kind of consciousness is the most relevant in our everyday lives. As a practical form of consciousness, it is connected with controlling action *per definitio*. It could be even maintained that acting morally is largely dependent on this ability. It merits the status of consciousness in that it describes well one's normal psychological states when one is able to report on and take responsibility for one's behavior. It is close to the juridical idea of whether one is conscious of what one has been doing. This way it is faithful to the etymology of the term consciousness in that it describes a form of mental functioning that includes an internal knowledge of itself and what is going on. One is conscious of what is happening. In this way, it is faithful to the layman intuition expressed by the transitivity principle. Also, as a more familiar kind of consciousness to the layman compared to the abstract phenomenality that is at stake in the hard problem of consciousness, it is likely that this kind of consciousness is the one the intuition of transitivity principle concerns.

I would even suggest that this is behind the intuition of higher-order theorists that there is more to being conscious than pure qualitiveness or phenomenality. They suggested that it being something like to undergo conscious experiences involves something more, namely a degree of awareness of one's mental states. I agree with them that self-awareness contributes to it being something like to undergo conscious experiences. Self-awareness makes conscious experiences more vividly conscious. What it is like to experience becomes more what it is like *for me* to be conscious. I can then be directed to objects with an "elevated form of self-awareness", to borrow the idea of Van Gulick (2004b).

Another way to characterize mentalized consciousness in contrast to reflexive consciousness would be in terms of transitive and intransitive consciousness. Introspection is transitive awareness of one's mental states: "I am conscious of seeing X". Mentalized consciousness is more of the intransitive kind, "I consciously see X", as far as awareness of mental states are concerned. Here, consciously means mentalized consciousness. I defined mentalized consciousness as dispositional self-consciousness, so a translation could be: "I potentially self-consciously see X". Other kinds of conscious states, like phenomenal experiences would be better characterized in the same manner: "I phenomenally see X". Mentalized consciousness is primarily a *world-directed* form of consciousness, which is informed and guided by the psychological dimension.

Of course, some kind of rudimentary higher-order representations are required in order for the capacity to mentalize to develop. In this way, mere dispositional higher-order theory cannot account for the concept of mentalized

consciousness but is required to include an account of the development of this capacity in relation to one's early social relationships. It continues to develop throughout one's life. Fonagy's idea that therapeutic change is partly enabled by the development of mentalization skills is one example of it. It is plausible that the most developed forms of reflexive consciousness build on the basis of the earlier mentalized consciousness.

5.4 The intuition of "givenness of consciousness"

Earlier, I criticized theories of phenomenal consciousness that are based on self-awareness. I granted them that they have to do with there being something it is like to have conscious states in that they determine a specific kind of consciousness, mentalized consciousness. The concept of mentalized consciousness is, however, based on a dispositional theory of self-awareness, so it is world-directed. An advantage to a theory would be also to be able to explain why there is the strong intuition that conscious states are ones we are aware of being in. In this section, I will attempt to deal with this by analyzing how this intuition of givenness of consciousness should be conceived of.

A coherent and satisfactory reconciliation would require, however, that a couple of demands that follow from the transitivity principle must be met. First, one must respect the intuitions of the layperson and those propounding the transitivity principle by at the very least attempting to explain the intuition supporting it. Moreover, our ability to reflect on our mental states seems to require some kind of givenness of consciousness. This possibility and the motivation for reflection should be explained. Zahavi accounted not only for these with his concept of pre-reflective self-awareness, but also for the *mineness* of mental states of consciousness. An alternative account must have answers to all these requirements.

I think that these requirements can be met if a certain ambiguity in the term 'givenness' in the expression "givenness of consciousness" is recognized. There are two senses in which "givenness" can be used. The first sense of "givenness" of consciousness is that consciousness is somehow *available* to itself. It is clear that something must be first available in some sense for us to be able to be aware of it. When it comes to mental states we can be conscious of only those mental states that are somehow available to us. But something can be given or present even when we manage not to be aware of it. The same goes for mental states. I shall henceforth call this sense "givenness as availability".

The point can also be formulated in information-theoretical terms familiar from the context of the contemporary first-order representational theories of consciousness (Dretske 1995). The givenness of consciousness itself in the sense of availability might mean that consciousness and its intentional object *carry information about* consciousness itself and the A-R distinction (Dretske, 1981; Aydede & Güzeldere, 2005). Dretske treats consciousness as an information-

using system. As such, it receives and seeks information-carrying signals and makes use of these signals by extracting pieces of information for further purposes. An important property of every information-carrying signal is that there is no limit to the amount of information that it can carry. Everything depends on the system's ability to extract the information from the signal. A signal can carry information about the system itself. Consciousness carries information not only about objects but also about itself. Thus, if consciousness has the proper capacities to extract the information, it may come to learn about itself. Whether or not it has the capacities, the information is always available. Consciousness is always given or present to itself in this way.

This sense is compatible with the principle of PE and the developmental theories. The account given by Fonagy et al., of how subsequent reflective capacities are enabled to develop, accords with the sense of givenness as availability. There could be other ways of accounting for the developmental process compatible with this sense. Givenness is here defined as the possibility or propensity to become aware of mental states, hence the demand of explaining the possibility of later reflective capacities does not pose any challenge for this sense of givenness.

The second sense of the givenness of consciousness is the *awareness* of the experience. This is the developed and derived form of the givenness of consciousness that comes about through conceptual development, the ability to extract and utilize the information that consciousness carries about itself, in either an implicit or an explicit manner. This sense, then, requires not only the first sense but also the existence of higher-order conceptual capacities to extract such information. Theories holding the transitivity principle require this sense of givenness; givenness as availability is not enough for one to be aware of one's states. Givenness is the result of the property of consciousness of having a dual content. Consciousness is directed to objects and to itself. Consciousness is given to itself as one of its intentional objects, albeit not the primary, intentional content⁷². I shall call this sense, where consciousness figures among the intentional objects of consciousness, "*givenness as awareness*".

Given these two senses of 'givenness', what about the phenomenological theories: how do they stand in relation to this equivocality of givenness? Although the phenomenological theories as described above were associated at least with a weak sense of the transitivity principle, they were also associated with some aspects of PE. I think the phenomenological theories can be made consistent with PE. In the following, I attempt to show how the distinction just made between the two senses of 'givenness' might alleviate the apparent conflict by reformulating some of the statements and transcendental arguments cited above from phenomenological theorists. In the citation above, Zahavi seemed explicitly to concur with the transitivity principle and givenness as awareness. Only a moment later, however, he describes Husserl's account in the

⁷² The situation is the same whether it is facts that are talked about instead of objects, or in other words, in addition to the primary intentional object, consciousness contains awareness of the fact that it is consciousness.

following, less demanding manner, which I think is akin to the sense of givenness as availability.

“According to Husserl, our acts are tacitly self-aware, but they are also accessible to reflection. They can be reflected on and thereby brought to our attention (Hua 4/248) [...] However, in order to explain the occurrence of reflection it is necessary that that which is to be disclosed and thematized must be (unthematically) present, otherwise there would be nothing to motivate and call forth the act of reflection. As Husserl points out, it is in the nature of reflection to grasp something that was already given prior to the grasping.” (Zahavi 2003, 88)

As noted earlier, however, the talk of (unthematical) *presence* as availability does not require awareness. Something can be present without our being aware of it. This formulation does not necessarily invoke awareness and the problems related to it. What is more important is that this passage concerns the very demands givenness should meet to allow further acts of reflection, the presence and the ability to motivate the act of reflection. Both of these go easily with the first sense of givenness, givenness as the availability of consciousness, without invoking problematic formulations associated with any form of awareness.

Therefore, I think that the transcendental arguments made by phenomenologists could be attenuated by replacing talk of awareness (or even acquaintance) by talk of availability. The arguments would still accomplish what they are supposed to. A transcendental argument can take many forms and the form should therefore be conceptually and empirically constrained. If the principle of psychic equivalence were taken into account, we should say, instead of claiming that we must be pre-reflectively *aware* of mental states, that they are *available* to us. In this vein, we could conceive of “living through the act” not necessarily by awareness of the act, as Husserl himself seems to have thought at times: “The appearing of the thing (the experience) is not the thing which appears [...] As belonging in a conscious connection, the appearing of the things is experienced by us, as belonging in the phenomenal world, things appear before us. The appearing of the things does not itself appear to us, we live through it” (Husserl 1970, 538). Here, he seems to suggest that living through the act is not the awareness, nor even the appearance of the experience. Rather, it is the *having* of the experience. Whether he really held this kind of view at other times is a question I leave to Husserl scholars. What matters is that this view is still available to phenomenologists and the one, to which in my view they should adhere.

A similar suggestion is, in fact, made by Amie Thomasson (Thomasson 2000; 2004 & 2005). Her one-level theory of consciousness, inspired by Husserl and Brentano, explicitly avoids defining consciousness in terms of awareness of mental states. Rather, such self-awareness is supervenient on phenomenal consciousness (cf. Thomas 2003). She wishes to retain the Brentanian-Husserlian idea of object-intentionality in virtue of living through the act, and thereby the distinction between the object and the experience of it, but not in the sense of being somehow aware of the act. In her adverbialist view, the difference between conscious and non-conscious mental states is that conscious states are

simply those by which we are *consciously* aware of things, those that we consciously live through. In her (2000) and (2004), however, she seems to commit herself to a reflexive content view much like that of Zahavi, but her (2005) seems to avoid this view. In this latter article, Thomasson seems to have become aware of this and develops her ideas in a strictly first-order or one-level form, without invoking any reflexive complex content (p. 12, see also p. 5). David W. Smith (1989, 102), on whose theory Thomasson leans to a large extent, explicitly dissociates the reflexive and phenomenal components of complex content, admitting the possibility of phenomenality without reflexivity. Thus, Thomasson and Smith seem to take the relation between phenomenality and awareness of it as contingent, not as necessary, as Zahavi does, although also in Smith's view the two virtually go hand in hand⁷³.

Thomasson also thinks that her view retains "a grain of truth" of the idea of TP, since in her view conscious mental states are *available* for reflective awareness (Thomasson, 2000, 205; 2005, 12). Her adverbialist view, then, constitutes a good account for availability to awareness of mental states. Any move to such an inner awareness, however, in her view requires a "cognitive transformation" where one abstains (in the manner of *ηποχη*) from worrying about the reality of objects present in awareness and concentrates on the way they appear. Her latest formulations appear to be in harmony with the first sense of givenness that I presented, givenness as availability, and thus hers seems the kind of view which other phenomenologically oriented theories of consciousness should opt for as well. Even with things put as she suggests they should, phenomenologists would nevertheless be fully warranted to continue their quest for the conditions of the self-manifestation of subjectivity.

A proponent of the phenomenological theory would object that my description of the principle of psychic equivalence and givenness as availability is exactly what they are trying to capture, and my exposition of its problemacy is due to terminological rather than substantive issues. I am very sensitive to this objection, especially since I think that this is a commonly encountered problem. However, I think the most fundamental issue, the cognateness of phenomenological theories with the transitivity principle – they are after all a form of self-awareness theories of consciousness – and the problems related to it is a substantial issue. I hope to have shown a way to avoid these problems through conceptual and terminological care.

In conclusion, I have criticized views of consciousness that require from a phenomenally conscious experience an awareness of itself as experience. Nevertheless, I tried to preserve a sense for givenness of consciousness that can survive this criticism. Consequently, I distinguished two meanings for 'givenness' of consciousness, givenness as awareness and givenness as availability, and proposed a dispositional view of givenness of consciousness along with preference for a first-order view of consciousness based on

⁷³ The reflexive content, if not adequate to characterize minimal phenomenal consciousness, would, nevertheless, be a good description of what I called "mentalized consciousness". In mentalized consciousness the phenomenal and reflexive contents go hand in hand.

developmental considerations. Availability or presence of consciousness without awareness of consciousness to itself meets the demand of higher-order views by allowing the possibility of reflection for what is a first-order view of consciousness. The implicit presence of self in the experience guarantees the mineness for the reflected states, as insisted on by phenomenologists. I do not pretend to have provided any knock-out arguments in favour of a first-order view of consciousness but I hope to have shown that the intuition of the transitivity principle is not as firmly established as its proponents uncritically take it to be.

5.5 Contenders and counter-arguments

The present view leans heavily on a dispositional higher-order theory in explaining mentalized consciousness and psychological introspection. Most contenders for the present view and their weaknesses have already been discussed. There are, however, one serious alternative and one line of criticism that have not been considered. An alternative account of higher-order representational abilities that does not address the concept of consciousness has most notably been made by Nichols and Stich (2002). Like the present proposal, their view is also inspired by empirical and clinical observations (mostly Frith 1992).

Nichols and Stich (2002) attack the dispositional higher-order theory by questioning the need for a higher-order thought system at all. For a theory account, having one and the same system to both detect and reason about mental states, a higher-order system is crucial. The problem with such a version is that it is highly implausible in its behaviorism (called “crazy version” by Nichols and Smith). In a view that postulates separate systems for detection and reasoning, it becomes questionable why a higher-order thought system would be needed, insofar as belief boxes already are present:

“What is needed is some source of information that would help a person form beliefs (typically true beliefs) about his own mental states. The obvious source of information would be the mental states themselves. So, on this version of the TT, the ToM has access to information provided by perception, information provided by background beliefs, and information about the representations contained in the Belief Box, the Desire Box [...] Now at this juncture one might wonder why the ToM is needed in this story. If the mechanism subserving self-awareness has access to information about the representations in the various attitude boxes, then ToM has no serious work to do. So why suppose that it is involved at all? That's a good question, we think. And it's also a good launching pad for our theory.” (Nichols and Stich 2002, 168–169)

This counter-argument seems to me problematic in the following ways. First, it does not suffice for self-*awareness* that mental states themselves carry information about themselves. Their carrying information about themselves is obvious. Nor does it suffice that the system has access to this information. What

more is needed is a means to extract and retain this information. This can be guaranteed only by a higher-order representation system. The model of Nichols and Stich can only guarantee givenness as availability, not givenness as awareness. What is symptomatic is that their own model indeed incorporates such a higher-order mechanism.

In their model, there is a separate monitoring mechanism in the “belief box”. It adds to every belief the prefix “I believe that”. The overall architecture is that there is an efference copy for each of our thoughts and perceptions, much like there is thought to exist for intentions and bodily movements (Frith 1992). This model, however, is multiply problematic. First, it is not thought that it provides a good theory of schizophrenia, which is the primary motivation for postulating such mechanisms (Zahavi 2000). Second, and more importantly, the existence of such an efference copy is taken to be redundant as long as it is possible to anticipate the outcome. Instead of sending efference copies, one simply anticipates an outcome and compares this to the afferent information (Gallagher 2005, ch. 8). Anticipation works by conceptual information, so a higher-order mechanism is well suited to play this role. Third, the view conflates first- and second-order representations. A mere belief box is usually meant to hold beliefs about the world. A disparate mechanism is required for reasoning about mental states; so much should be clear from the results of the false belief experiments.

Finally, and most importantly, the model of Nichols & Stich is intended to explain something like mentalized consciousness. The development of a monitoring capacity remains unclear, although nothing other than some sort of genetic determination could explain why the capacity should remain as contrived as it is; there is only a registration of beliefs, nothing more. The concept of mentalized consciousness suggested here does, however, differ from this in that it consists in capacities of self-regulation. The self-regulation allowed by register-monitors does not extend this far without further conditions, ones that would blur the border-zone with this view and the higher-order theories. Higher-order theories also fair better than mere monitoring mechanisms in explaining how we can enjoy more developed modes of self-monitoring, those of introspection and reflexive consciousness as described above. In addition, the possibility of consciousness without self-awareness remains unclear. These features all support a dispositional higher-order-theory of consciousness.

Some may still wonder whether my way of defining higher-order representational consciousness suffers from the problems of act-object theories, because of adherence to the modality-mode model of Smith (1989), or adverbialism, since that is where I situated consciousness and whence I took it to be the object of introspection.

Revonsuo denies that either of these aspects (acts or adverbialism) really exists in consciousness. I agree with this in that I do not take the object to be something external but as a part of consciousness. “Patterns of phenomenal experience simply come about as inherently self-presenting; they are neither acts or objects (or, if you prefer to put it in this way, they are both the act and

the object at the same time)" (Revonsuo 2006, 132). Modes (objects) and (modalities) are inextricably bound to each other. The same could be said about Revonsuo's own view of consciousness as virtual reality: it is constituted by "phenomenal space". This view describes the contents of consciousness, nevertheless, as things in an, albeit virtual, world. Experientially, these contents are something other than myself, and in that sense acquire their otherness by becoming contrasted to a self. Further, Revonsuo claims that there might not be any self-model present in experiences relying on introspective reports of dreamers who claim to have had totally selfless experiences. However, that there could be some implicit self-structures governing these contents that do not appear to the subject as thematic contents, which would be compatible with their introspective reports. Be that as it may, I am prepared to allow wholly selfless experiences. This would then amount to an experience of "buzzing blooming confusion" as described by James (1890, 462) or another sense of confusion: the original "undifferentiatedness" described by the psychodynamic theories. It could be that reflecting on such an experience would not be experienced as mine. Nonetheless, it would amount to self-awareness from an external perspective: the reflecting subject would be aware of his own states.

The traditional problems of adverbialism do not pose a threat to the present view. The point of traditional adverbialism is to dispense with the object, and only do with the act. Therefore, it analyzes contents as adverbial, and it is what the traditional counter-arguments leveled at adverbialism are concerned with. In the present view, contents are not analyzed adverbially, only manners of representation, in order to give plausibility to the possibility of introspection. In introspection, one abstains from the truthfulness of the first-order states - similarly as in standard adverbialism. However, one then takes the content of introspection to be true and the accuracy of the second-order consciousness can then be assessed.

6 CONCLUDING REMARKS

The present discussion of the higher-order theories had a two-fold purpose. First, it attempted to undermine the much-wanted success of the higher-order theories in the task of explaining the hard problem of consciousness, how conscious phenomenality can arise from its cognitive and neural underpinnings. I argued that invoking higher-order representations can hardly help in this task, since while conforming to some intuitions, doing so conflicts with others. HO theories lack empirical plausibility, and worst of all, they do not offer an account for why they should work in the first place. This is simply assumed not explained. Yet, conscious phenomenality should somehow be explained. In many cases, the intuition behind HO theories breaks down, and we tend to think that infants and animals undergo phenomenal experiences without being able to be aware of such phenomenal states. It is more plausible to think that such awareness of one's phenomenal states is developed to improve self-regulation, and it is not necessarily the source of the phenomenality of experiences.

Self-regulation and self-awareness, however, seem like something that laypersons most often refer to by the term 'consciousness', and this may be the source of the intuition that being aware of one's phenomenal states is essential for consciousness. On this basis, I suggested that awareness of one's mental states or higher-order representations are essential requirements for a distinct kind of consciousness, "mentalized consciousness", which constitutes another, more developed meaning of there being something it is like to be conscious. The concept of "mentalized consciousness" corresponds better to the intuition behind the HO theories of consciousness involving self-awareness, and is thus informative about the relation between consciousness and self-awareness. By mentalized consciousness I mean a world-directed consciousness that is informed and guided by psychological information and the ability to regulate oneself on the basis of this information, an ability which is developed little by little, primarily in one's earliest interactions with one's caregivers and secondarily with other important people in one's personal history. It was also described in terms used in the clinical and cognitive sciences, since it is implicit and cannot be described by mere folk-psychological terms. I tried to build an

empirically informed model of the concept at multiple levels and along a developmental dimension. In binding together descriptions at multiple levels, it is an example of the strategy based on co-evolution and semantic monism that I described in the first-section. I contend that the concept of mentalized consciousness is a result of a coevolutionary strategy: conceptual and scientific evidence guide the description of the kind of consciousness that involves self-awareness. This way, it gains plausibility over the abstract kind of self-awareness described by the classic higher-order theories. Some suggestions were made on the way as to how *phenomenality* should be treated in the same explanatory framework (though see Revonsuo 2006 for a more thorough-going proposal).

Higher-order theories are also useful in differentiating different forms of self-awareness. They cannot explain the most primitive form of self-awareness, the presence of a self or the first-person form of mental states, conscious or non-conscious, the phenomenon called “pre-reflective self-awareness” by phenomenologists, but rather presuppose it. My view is that the concept of pre-reflective self-awareness should not be spelled out as awareness of one’s own mental states but rather as indexicality, presence of invariant self-aspect in the mental states. Indexicality brought about by different self-systems suffices to explain how higher-order theories yield self-knowledge, not knowledge about the world. However, the HO theories are useful in elucidating the nature and forms of metarepresentational activity: introspection, reflexive consciousness and mentalized consciousness. Both Block (1995) and Chalmers (1996) have made this remark about this possible application of higher-order theories. It has not been carried out previously, and this has been my task in this work. The theories and the lively debate between the proponents of its different forms are also very useful in analyzing the nature and underpinnings of higher-order theories in general.

In the same vein, Rosenthal maintains that phenomenal consciousness holds no special status, since it does not really capture the term consciousness, which involves a degree of self-awareness. Other major higher-order theorists, in fact, concur with this, and it is legitimate to say that they are explaining some other meaning of consciousness than simple phenomenality. Many rebuttals levelled at higher-order theories that stem from the implausibility of the HO theories in explaining phenomenality are misguided from this perspective. It is also true that higher-order theories do hold some intuitive plausibility. It remains, however, unclear what that intuition is based on, and what kind of consciousness HO theories are supposed to explain. My suggestion is that what they could be used to explain is rather the kind of consciousness that I call mentalized consciousness.

Block thinks that we should not reserve special status for any of the concepts of phenomenal, access, monitoring and self-consciousness (Block 2002): they all deserve the status of consciousness. I further distinguished between different forms of self-consciousness, reflexive and mentalized consciousness. The conflict between these higher-order and first-order views evaporates once mentalized consciousness is introduced to the picture. Through

this concept it becomes intelligible that creatures without higher-order representations could have phenomenal experiences. Nonetheless, it retains the intuition of control and self-awareness in another form of consciousness characteristic of human beings. Previously, this aspect was confined to reflexive consciousness. The concept of mentalized consciousness includes both aspects, the emphatic outwardness of normal consciousness, yet retaining an element of inwardness in consciousness and incorporating more organization and unity with an “enhanced aspect of metaintentionality” (Van Gulick 2004b, 80).

In addition to my suggestion for a mentalized kind of consciousness, the debate over the higher-order theories of consciousness lends itself readily to the debate over the nature of introspection. Different positions yield different accounts of introspection. Some arguments that pertain specifically to consciousness can be dismissed, but many of the arguments remain valid. My contention was that the best theory is achieved if we deny the opposition between higher-order thought and higher-order perception. Introspective access is of neither of these kinds, it is something that has no word in the folk-psychological vocabulary. Instead, its nature is best described by the vocabularies of cognitive science, by the sort of access that is spelled out by the global availability theory as access to working memory. It is a sort of attention that is guided by specific concepts, psychological and phenomenal concepts, as spelled out by an informational theory of Aydede & Güzeldere (2005).

Dretske’s conception of introspection was dismissed, since the object of phenomenal concepts is not the object in the world, but rather the way the object seems to be presented – the way consciousness is “*as of*” an object. The nature of “*as of*” is the very same transparency that first-order theorists appeal to in connection with phenomenal consciousness (see ch. 3). This transparency is the way how consciousness of being as of objects. Such a view adheres to the view that consciousness has narrow content. When the fact that introspection is about the object-directed nature of *consciousness*, not the object consciousness is about, is conceded, it follows that a first-order representational theory cannot speak about introspection without invoking higher-order representations of some form. My own view is that higher-order consciousness takes the modality and mode of presentation as its object. Modalities correspond to types of mental states, and mode of presentation is phenomenologically determined narrow content that is assessable for accuracy (cf. Siewert 1998). In introspection, however, this assessability is not at stake, so introspection is in this sense more like “bracketing” as described by phenomenologists (or cognitive transformation; see Thomasson 2004).

I think that these are all sensible suggestions that help to avoid most of the traditional problems of the (especially second-order) representational theories. The first demand for a theory of higher-order representational theory – elucidating the nature of the privileged access to one’s mental states – can be formulated in the spirit of global access theory rather than higher-order thought or perception account. The higher-order views commit pushing a form of external intentionality into the mind. The second and the third demands were explaining who is aware of experiences and how self-awareness is possible in

the first place. These are secured by the view of self as the deepest *context* of experience, both spatially and at the level of goals, attitudes etc. The self is nothing over and above the structure of the egocentric space with the representation of body in the middle of it. This kind of indexical structure of consciousness can serve as a minimal form of self-familiarity (the same self-structures are present in the succession of mental states with distinct varying contents). By this indexical self-familiarity it can be explained why reflection constitute self-awareness. It was also maintained that this entails the peculiar semantic properties of first-person thoughts.

Although not explored here in detail, the results of this study and the concept of mentalized consciousness in particular hold a promise of possible application and further studies. One important issue to which HO theories are taken to pertain is the question of personhood. Personhood has been connected to self-awareness (Frankfurt 1971; Dennett 1975; Wilkes 1987; Baker 2000; Rosenthal 2002). The concept of mentalized consciousness could bring a new dimension to this debate. The kind of self-awareness considered by these authors is second-order desires (Frankfurt), conceiving of oneself as oneself (Wilkes & Baker) or higher-order thoughts (Dennett, Wilkes & Rosenthal). In my view, all of these are defective in that they are either too restrictive, especially the second-order desires of Frankfurt or they are too abstract and formal like the ability to have higher-order thoughts or conceiving of oneself as oneself.

As far as characterizing personhood is concerned, mentalized consciousness fares better than these views in that it is more practical and context-bound. This, naturally, corresponds to interactions between real persons. Mentalized consciousness is characteristic of persons, since self-regulative capacities as explained in terms of mentalization ability and history constitute largely the roles or "masks" that we wear in social interaction (word *persona* originally meaning mask and later a role). In juridical contexts, consciousness means most probably this sort of consciousness, i.e. in whether one is conscious of one's acts or not. If one loses "it" (perhaps meaning one's mentalization ability), one is held not responsible for one's actions as one has no control over them. One is, in one sense of the term, not conscious of one's actions. As a practical form of consciousness, mentalized consciousness can readily explain the normal conscious actions better than the other proposed forms of consciousness. It thus conforms well to what people typically think about persons.

Questions about personal identity typically look for a definition of personhood and for necessary and sufficient conditions. I do not think that mentalized consciousness can be either, but neither do I think that seeking definitions about personhood is fruitful in the sense that it neglects information about typicality. Initial descriptions about something do, however, start from typical instances and mentalized consciousness might be important there as a starting point (cf. Wilkes 1987; Dennett 1975). I would not be surprised, however, if something like mentalized consciousness would be fruitful in determining the persistence conditions of persons. In my view, the most

important aspect about individuation and the persistence of objects is how they maintain themselves. What we take to be material things are ones that are relatively stable as compared to other configurations of materia. Typically, they are self-maintaining and self-regulating either by virtue of their physical, biological or psychological properties. Self-regulation capacities thus determine how long a person persists as a person. Other theories of personhood based on self-awareness cannot as readily explain this aspect, since they are too formal, tying personhood to a presence of a single abstract capacity. Mere ability to reflect on one's mental states or conceive of oneself as oneself cannot guarantee self-regulation. Excessive self-awareness can even compromise self-regulation, as might be the case in schizophrenia (Zahavi 2000) or eating disorders (Fonagy et al. 2002). A more practical and context-bound view is required to handle these cases. Probably something like mentalized consciousness is what other self-awareness theories of personhood try to mean. If this is the case, they under-describe the developmental and practical dimensions of self-awareness. A more empirically informed account is needed in any case.

If the concept of mentalized consciousness were to prove successful in elucidating the nature of personhood, it might bear fruit in the branches of ethics, social philosophy and even clinical psychology. The concept of mentalization is, of course, already an established part of the last-mentioned, but the dimension of consciousness might bring a new flavour into that field, too. There is as yet little clinical use of the concept of consciousness except in neuropsychology, but psychopathology could be an interesting area of application, and the concept of mentalized consciousness could open a new perspective on it.

Another field where HO theories have been applied is the question of animal consciousness. In a widely accepted view, animals have phenomenal consciousness or even rudimentary forms of access consciousness. On the grounds of the HO theories, this has been denied (Carruthers 1989; 1998b) or the HO theorists have lamented their views to accommodate this intuition, thereby rendering themselves in conflict with other intuitions, and compelled to postulate higher-order representations to primitive creatures (Rosenthal 1997b; Gennaro 2004). My view, however, offers a means to avoid conflicts with these general intuitions. Animals in general do not have mentalized consciousness in the same way that we do because of the lack of adequate attachment relations that make mentalized consciousness possible, although phenomenal and access consciousness could be abundant in the animal kingdom. It seems entirely possible that at least some of the human ape species could have a sort of mentalized consciousness. However, occurrent, focal introspection or engaging in meditations of reflexive consciousness seems beyond the cognitive capacities of non-human animal species.

If the distinctions made between different forms of consciousness and self-awareness correspond to real phenomena, an interesting field of application would be the cognitive and neural sciences. Different concepts of consciousness, phenomenal and access consciousness, have already proven to be of use for scientific experiments. The dimensions of reflexive consciousness, introspection

and mentalized consciousness could also prove useful in devising scientific experiments and in interpretations of their results. It also would be a good place to test whether they are real or not.

Last but not least, the considerations of this study bear a more general relevance to the empirical sciences. I claimed that all psychological studies must start from the pre-theoretical conception of mind that is given to oneself in the first-person experience. In fact, our conception of our minds would be completely different without the first-person access. If the view of multi-level theories of mind is correct, this means that our first-person access constrains even the neural sciences. The findings in neural sciences are ultimately incorporated to the multi-level model, through which the findings get their meaning. Therefore, the first-person methodologies are necessary for neural sciences of the mind. My first and foremost hope is that the work reported here is of use to those who use introspective reports in experimenting or study introspection by empirical methods, experimental or introspective, and have searched for a theoretically articulate account of the nature of introspection. Methodological issues about introspection cannot be settled without an account of what introspection is. The same account may be helpful in deciding the normative issues surrounding introspection: how can we develop as better introspectors, or gain a deeper self-awareness in general? The higher-order theories can provide us with such an account.

YHTEENVETO

Tajunnallisuus subjektiivisesti ilmenevänä kokemuksena muodostaa mielenfilosofian merkittävimmän ongelman – ja kenties yhden luonnontieteen suurimmista haasteista: kuinka on mahdollista, että aivojen toimintaan näyttäisi liittyvän subjektiivinen kokemus? Miksi minusta, biologisesta organismista tuntuu joltain olla minä, tietoinen olento? Säilyttääksemme luonnontieteellisen maailmankatsomuksen meidän olisi kyettävä selittämään subjektiivisesti ilmenevän kokemuksen tason ilmiöt tieteellisen psykologian, kuten kognitio- tai neurotieteen tason avulla.

Monet mielenfilosofit ovat päätyneet naturalismin kannalta pessimistiseen lopputulokseen subjektiivisen kokemuksen luonnontieteellisen selittämisen suhteen: se on mahdotonta. Niin irrallaan ymmärryksemme subjektiivisesti ilmenevän tajunnallisen kokemuksen tasosta on tieteiden tasosta. Perusteluksi on esitetty ajatusskenaarioita, joissa kehoitetaan kuvittelemaan kaksi kognitiivisilta ja aivofysiologisilta ominaisuuksiltaan identtistä olentoa. Antinaturalistit väittävät, että minkäänlaista käsitteellistä ristiriitaa ei muodostuisi, vaikka vain toisella näillä olennoilla olisi toisistaan poikkeavat tajunnallinen kokemusmaailma tai jos se toiselta puuttuisi täysin. Tajunnallisuus olisi tämän skenaarion mukaan täysin oma todellisuuden lajinsa, joka ei aukottomasti määräydy fyysikaalisen todellisuuden kautta eikä olisi siten ymmärrettävissä luonnontieteiden avulla. Minkäänlainen tieteellinen tieto ei voisi tuoda ratkaisevaa edistystä tajunnallisuuden ymmärtämisen kannalta – tajunnallisuuden ongelma on siksi ratkaisematon. Monille se olisi riittävä syy luonnontieteellisestä maailmankuvasta luopumiseen.

Luonnontieteellisesti orientoituneet tajunnantutkijat ovat karsastaneet tällaisia nojatuolista esitettyjä filosofisia käsitteellisiä argumentteja. Heidän mukaansa kokemuksen taso on mahdollista liittää yhteen kognitiivisten tieteiden tai neurotieteiden tiedon kanssa. Tajunta olisi tällöin luonnonilmiö siinä missä muutkin. Yksi ehdotus tajunnan ongelman ratkaisemiseksi on ollut selittää se tietoisuuden representationaalisen funktion avulla: tajunnansisällöt edustavat asioita maailmassa ja voivat erehtyä niiden suhteen. Kenties kokemuksellisuus on sama kuin sen evoluutiossa kehittynyt tehtävä edustaa asioita maailmassa. Kokemuksemme sisällöthän ovat maailman sisältöjä: kuulemani puhe on keskustelukumppanini puhetta, ulkoinen maailma avautuu näköaistini kautta ja tunnen maanpinnan jalkojeni alla. Jos tajunnallisuus on yhtä kuin kyky edustaa maailman asioita, niin voimme selittää kokemuksen subjektiivisesti ilmenevän aspektin. Meidän tarvitsee vain selittää tajunnan merkityksellisyys. Siitä puolestaan on olemassa lupaavia teorioita, kuten biologiset evolutiiviseen historiaan perustuvat teoriat.

Maailmasuuntautuneisuuden, kognitiivisen tai aivotoiminnan kautta selittävä teoria on muodoltaan ensimmäisen kertaluvun teoria, joiden mukaan tajunnallisuus koostuu vain kokemuksellisesta tai siihen liittyvästä merkityksellisyyssaspektista. Kilpailevana ehdotuksena tälle on esitetty joukko korkeamman

kertaluvun teorioita, joiden mukaan ensimmäisen asteen teoria ei voi olla riittävä: se ei kykene erottamaan ei-tietoista mielen toimintaa perustavanlaatuisesti tietoisesta toiminnasta. Se ei myöskään pysty itsensä sisällä erottamaan mieltä ja maailmaa toisistaan. Se, mikä puuttuu, on tietynlainen itsetietoisuus, kyky olla tietoinen omista tietoisuuden tiloista. Korkeamman kertaluvun tietoisuus-teorioiden ytimen muodostaa yleinen intuitio, että kokemuksellisesti tietoisia mielen tiloja ovat vain ne, joista olemme jollain tavalla tietoisia. Tätä ajatusta korkeamman kertaluvun tietoisuusteoreetikot kutsuvat transitiivisuusperiaatteeksi (engl. *transitivity principle*). Vain ollessamme tietoisia kokemuksistamme meistä tuntuu joltain kokea niitä. Muu mielen toiminta, josta emme ole tietoisia, on eri tavoin ei-tietoista toimintaa. Sellainen mielen toiminta ei tunnu meistä miltään. Korkeamman kertaluvun tietoisuusteorioiden ratkaisevat tajunnallisuuden ongelman tämän intuition avulla. Jos voimme kognitiivisen psykologian termin selittää itsetietoisuuden omista mielen tiloista, niin voisimme selittää subjektiivisesti ilmenevän tajunnallisuuden.

Itsetietoisuudesta tietoisuutena omista mielen tiloista on olemassa joukko teorioita. Itsetietoisuudella on muutama erityispiirre, jotka teorian pystyttävä ottamaan huomioon ja joiden avulla teorioita voi arvioida. Ensiksi, meillä näyttäisi olevan jonkinlainen yksityinen pääsy omiin mielen tiloihimme: voimme tietää ja tarkkailla niitä tavalla, joka ei ole muille mahdollinen. Lisäksi tietoisuus omista mielen tiloista näyttää sisältävän tiedon, että ne ovat juuri minun mielentilojani. Näiden asioiden ohella on kyettävä selittämään, millä tavalla psykologista tasoa koskevaa tietoa, johon tieto omista mielen tiloistamme kuuluu, voidaan säilyttää, käsitellä ja käyttää hyväksi.

Korkeamman kertaluvun teorioita on kaksi klassista muotoa, jotka poikkeavat erityisesti sen mukaan, millainen pääsymme mielen tiloihimme niiden mukaan on. Ensimmäisen mukaan meillä on jonkinlainen sisäinen aisti, jonka avulla kykenemme tarkkailemaan omia mielentilojamme. Kanta on virinnyt jo vuosituhansia sitten, ja sen intuitiivinen vetoavuus on vahva. Yleisen ilmaisun mukaisesti monesta tuntuu, että kykenee ”mielen sisäisen silmän” avulla tarkkailemaan oman mielensä sisältöjä samoin kuin aistinvaraisesti voi tarkkailla ulkoisen maailman sisältöjä. Teorialle ei kuitenkaan ole onnistuttu esittämään sitä tukevaa tieteellistä todistusaineistoa, ja sisäisen aistin hermostolliset mekanismit ovat hämärän peitossa. Toinen klassinen kanta luottaa puhtaasti käsitteellisiin kykyihin: voimme tulla tietoiseksi mielentiloistamme ajattelemalla niitä. Sisäisen aistin puuttuminen ei ole ongelma tälle kannalle, ja lienee selvää, että ainakin meillä on kyky kohdistaa ajattelumme koskemaan omaan ajattelutoimintaamme. Se ei kuitenkaan vastaa yhtä hyvin itsetarkkailukykyämme koskevaa ”sisäisen silmän” intuitiota.

Molempien klassisten kantojen edustajat ovat kyenneet osoittamaan vastapuolen kannassa heikkouksia, ja molemmilla on vahvuutensa. Väitöskirjassa muotoilen näitä kantoja paremmin nykyiseen kognitiivisiin, psykologiseen ja neurotieteeseen sopivaa korkeamman kertaluvun teoriaa, joka kykenee hyödyntämään molempien vahvuuksia ja täyttämään samalla edellä kuvatut kriteerit hyvälle itsetietoisuusteorialle.

Kuvatuissa klassissa kannoissa tietoisuus mielen tiloista on kohteena olevista mielen tiloista erillinen akti. Lisäksi on olemassa kolmas itsetietoisuuteen perustuva teoria; sen mukaan jokaiseen mielen tilaan sisältyy tietoisuus itsestään. Se vastaa parhaiten intuitiota siitä, että tietoihin mielen tiloihin liittyy tietoisuus siitä itsestään. Erityisesti fenomenologisten teorioiden piirissä näkemys on tullut suosituksi. Fenomenologisessa perinteessä itsetietoisuutta ilmiönä ja sen ehtoja on kuvattu laajasti, erityisesti mielen tilojen kuulumista minulle. Fenomenologit ovat esittäneet, että itsetietoisuuden tätä ominaisuutta ei voi kuvata tai selittää tieteellisin termein joutumatta käsitteellisiin ongelmiin. Se on siksi ehdotettu tajunnallisuuden primitiiviseksi ominaisuudeksi ja välttämättömäksi ehdoksi. Näkemys puoltaa transitiivisuusperiaatteen ohella itsetietoisuuden pitämistä tietoisuuden ja tajunnallisuuden sisäisenä ominaisuutena. Pyrin työssä näyttämään, että fenomenologien johtopäätökset ovat turhan pitkälle vietyjä. Tietoisuuden ei tarvitse sisältää itsetietoisuutta tietoisuuden tiloina, vaan on riittävää, että se on indeksikaalinen eli sisältää jonkinlaisen itseysaspektin. Tietoisuus sisältää näin jaon organismiin itseensä ja kaikkeen muuhun. Itseysaspektiin kuuluvat suhteellisen pysyvät "itsejärjestelmät", kuten oman kehon ja mielen representaatiot, niihin liittyvät toiminta- ja tunnetaipumukset sekä itseä koskevat muistot ja tavoitteet. Näiden suhteellisesti pysyvien taustasisältöjen muodostaman itseyden kontrastoituessa vaihtuviin tietoisuuden sisältöihin voidaan ymmärtää jatkuvan identiteetin tunne sekä tunne tietoisuuden tapahtumien kuulumisesta yhdelle ja samalle subjektille - minulle.

Arvioni korkeamman kertaluvun teorioiden strategiasta selittää kokemuksellisesti ilmenevä tajunnallisuus itsetietoisuuden kautta on, että se ei intuitiivisesta vetovoimastaan huolimatta voi onnistua. Sen tärkein ongelma on, että strategian toimivuutta tai transitiivisuusperiaatetta ei varsinaisesti selitetä. Miksi itsetietoisuudesta pitäisi seurata itsetietoisuus? Näyttäähän olevan runsaasti tapauksia, missä itsetietoisuudesta ei seuraa tajunnallinen kokemuksellisuus. Ja eikö usein pikemmin ole niin, että intensiivisempi kokemus toteutuu, kun on niin vahvasti maailman sisältöihin uppoutuneena, että unohtaa itsensä ja omat mielentilansa? Reflektoidessa omaa mielentilaansa tuo mielen tila on pikemmin haalistunut versio alkuperäisestä kuin se alkuperäinen maailmasuuntautunut elävä kokemus. Eikö mielen tilan ennemmin pitäisi olla alun perin tietoinen ja kokemuksellinen, jotta siitä ylipäänsä voi tulla tietoiseksi? Entäpä olennot, joilla ei ole kykyä itsetietoisuuteen, kuten lapset ja eläimet? Voiko hiiri olla tietoinen omista mielen tiloistaan? Jos ei voi, eikö sillä esiintyisi subjektiivista kokemusmaailmaa lainkaan? Suurimman osan mielestä esiintyy. Yhdessä näillä näkemyksillä on mielestäni enemmän intuitiivista vetovoimaa kuin korkeamman kertaluvun tietoisusteorioiden perusperiaatteella. Korkeamman kertaluvun teorioita koskeva intuitio tuntuu pätevän vai osittain. Lisäksi, kuten todettua, niillä on ollut vaikeuksia esittää niitä puoltavaa tieteellistä kokemusperäistä todistusaineistoa, mikä vaikeuttaa niiden asemaa entisestään.

Korkeamman kertaluvun teorioilla ja niiden ytimen muodostava intuitio tuntuu kuitenkin pätevän osittain ja sillä on paljon vetovoimaa. Korkeamman kertaluvun teorialat ansaitsevat siksi tarkemman analyysin ja tarkastelun niiden

täydestä potentiaalista tajunnantutkimuksessa on, eivätkä kriitikot ole tätä pyrkineet selvittämään yksityiskohtaisesti. Korkeamman kertaluvun teorioiden edustajat puolestaan väittävät, ettei heidän tarkoituksensa olekaan selittää pelkkää subjektiivisesti ilmenevää kokemuksellisuutta. Heidän mukaansa se, miltä tuntuu olla tietoinen olento, sisältää muutakin – erityisesti niiden kuvaaman itsetietoisuuden. He eivät kuitenkaan ole tarkasti kuvanneet, minkälaista tajunnallisuuden lajia he pyrkivät selittämään. Tässä työssä pyrim korjaamaan tämän puutteen kuvaamalla omia mielentiloja koskevaa tietoa ja itsesäätelyä koskevien kehityspsykologian teorioiden ja tutkimustulosten avulla tajunnallisuuden muodon, jonka korkeamman kertaluvun tietoisuusteorioiden voisivat selittää – tai pikemmin tajunnallisuuden muodon, joka selittää korkeamman kertaluvun teorioita tukevan intuition.

Tajunnallisuudelle ja tietoisuudelle on erotettu useita merkityksiä. Ehdotan, että tajunnallisuudesta tulisi erottaa itsetietoisuuden suhteen kaksi lajia, joista yksi, edellä kuvattu kokemuksellinen tajunnallisuus, ei vaadi tietoisuutta siitä itsestään. Mitä primitiivisempää tajunnallisuuden tasoa kohti mennään sitä pahemmin korkeamman kertaluvun teorioiden peruspilarin muodostava intuitio murenee. Siksi alkeellisempien tajunnallisuuden muotojen selittämiseen soveltuu todennäköisesti paremmin ensimmäisen kertaluvun teoria. Toiseksi on olemassa tajunnallisuuden tai tietoisuuden merkitys, joka sisältää itsetietoisuuden elementin. Kutsumme usein, esimerkiksi juridisissa konteksteissa tietoisuudeksi sitä, kun joku toimii tietoisesti ja kykenee psykologisen tason tiedon avulla säätelemään omaa toimintaansa: onko esimerkiksi rikos tehty tietoisesti vai ei. Käytämme termiä tietoisuus kenties useimmin tässä merkityksessä, mutta emme silloin viittaa ainoastaan subjektiivisesti ilmenevään kokemukselliseen puoleen mielestä, vaan johonkin kehittyneempään tietoisuuden muotoon. Itsesäätely on tämän kehittyneen tietoisuuden muodon suhteen tärkeässä asemassa. Kutsun tätä tietoisuutta ”mentalisoiduksi tietoisuudeksi” viitaten psykodynaamisesti virittyneeseen kehityspsykologian keskusteluissa esintyvään mentalisaation käsitteeseen. Psykodynaamisen mentalisaatiokäsitteen mukaan edellytykset kyvyille ymmärtää omaa ja muiden mieltä ja toimintaa kehittyvät varhaisessa hoivasuhteessa. Lapsi tulee vähitellen tietoiseksi sisäisistä tiloistaan ja oppii säätelemään niitä, jos hoivaaja tarjoaa siihen mallin oikeanlaisen tunnistamisen ja niihin reagoimisen kautta. Tämän mallin sisäistäminen tarjoaa ensimmäisen tavan tunnistaa ja säädellä omia sisäisiä tilojaan. Oikeansuuntainen kehitys mahdollistaa vähitellen kyvyn säädellä yhä itsenäisemmin omia sisäisiä tiloja, omaa toimintaa ja ajattelua mutta vinoumat hoivasuhteessa voivat johtaa paikallisiin häiriöihin itsesäätelyssä. Mentalisaatioteoreetikoiden mukaan joihinkin psykopatologioihin sisältyvä kyvyttömyys itsesäätelyyn, tehdä tietoisesti ja siten syyntakeettomasti jotain, on seurausta puutteista varhaisten kehitysvaiheiden hoivasuhteessa. Psykoterapian yksi tavoite on mentalisaatioteoreetikoiden mukaan tällaisen itsetietoisuuden kehittäminen ja saada ihminen siten toimimaan entistä tietoisemmin ja harkitsevammin. Suotuisassa kehityksessä itsesäätelykyvyt automatisoituvat kehityshistorian saatossa ja sisältävät lopulta suuren määrän käytännöllisiä itsesäätelykykyjä ja mahdollisuuden te-

mentisoida mielen tilamme tarvittaessa. Tällainen psykologisen informaation käyttämisen mahdollisuus voisi selittää, miksi ajattelemme tietoisien mielen tilojen sisältävän tietoisuuden niistä.

Mentalisoitu tietoisuus on kuitenkin maailmasuuntautunutta ja siten käsitteenä sallii intuition intensiivisen asioihin uppoutumisen kokemuksen itsetietoisuuden kustannuksella. Mentalisoidun tietoisuuden käsitteen avulla voisi myös selittää ihmisen tajunnallisuuden ja eläinten vastaavan laadullisia eroja, eikä mentalisoidun tietoisuuden käsite kärsi korkeamman kertaluvun ongelmista puhtaan kokemuksellisen tajunnallisuuden selittämisessä. Samalla se voi selittää korkeamman kertaluvun tietoisuusteorioiden intuitiivisen vetovoiman. Korkeamman asteen tietoisuusteorioita voi siten hyödyntää tämänlaisen tietoisuuden selittämisessä, vaikka ne epäonnistuisivat puhtaan kokemuksellisen tajunnallisuuden selittämisessä. Työn kahdessa viimeisessä luvussa kuvaan yksityiskohtaisesti, kuinka korkeamman asteen tietoisuusteorioita voi hyödyntää mentalisoidun tietoisuuden ja tietoisien oman mielen itsetarkkailukyvyyn, introspektion, ja niiden erityispiirteiden selittämiseen. Näytän myös, että tämä käsite voi tarjota korkeamman kertaluvun teorioita parempia mahdollisuuksia käsitellä tajunnan ongelman lisäksi muita filosofisia ongelmia, joihin niitä on sovellettu, kuten persoonuutta koskevaa ongelmaa. Korkeamman kertaluvun teoreetikot ovat ehdottaneet, että tietoisuus omista mielen tiloista erottaa persoonat muista olennoista. Ehdotan, että korkeamman kertaluvun teorioiden kuvaaman abstraktin itsetietoisuuden sijaan persoonina pitämiemme olentojen tyypillinen itsetietoisuuden muoto on käytännöllinen mentalisoitu tietoisuus.

REFERENCES

- Abu-Akel, A. 2003. A neurobiological mapping of theory of mind. *Brain Research Reviews*, 43, 29–40.
- Adams, F. & Dietrich, L. 2004. Swampman's revenge: Squabbles among the representationalists. *Philosophical Psychology*, 17, 323–340.
- Adams, W. 2000. Introspectionism reconsidered. Delivered at "Towards a Science of Consciousness (Tucson 2000)", April 10–15, 2000, Tucson Convention Center, Tucson, AZ.
- Ainsworth, M., Blehar, M., Waters, E., & Wall, S. 1978. *Patterns of attachment*. Hillsdale, NJ: Erlbaum.
- Alston, W. 1971. Varieties of privileged access. *American Philosophical Quarterly*, 8, 223–241.
- Anscombe, E. 1994. The first person, In Cassam, Q. (Ed.) *Self-knowledge*. Oxford, England: Oxford University Press, 142–159.
- Aristotle 1986. *Περὶ ψυχῆ* (On the Soul). London: Richard Clay Ltd.
- Armstrong, D. 1968. *A materialist theory of the mind*. London: Routledge.
- Armstrong, D. 1984. Consciousness and causality. In D.M. Armstrong & N. Malcolm 1984. *Consciousness and causality: A debate on the nature of mind*. Blackwell.
- Armstrong, D. 1997. What is consciousness? In N. Block, O. J. Flanagan, & G. Güzeldere (Eds.) *The nature of consciousness: Philosophical debates*. Cambridge, MA: MIT Press, 721–728.
- Arstila, V. 2005. *The paradox of colors*. Reports from the Department of Philosophy, 15. Turku: University of Turku. PhD Thesis.
- Astington, J, & Gopnik, A. 1991. Developing understanding of desire and intention. In Whiten, A. (Ed.) *Natural theories of mind: Evolution, development and simulation of everyday mindreading*. New York: Blackwell, 39–50.
- Austin, J. 1962. *How to do things with words*. Cambridge, MA: Harvard University Press.
- Aydede, M. 2001. Naturalism, introspection, and direct realism about pain. *Consciousness and Emotion*, 2, 29–73.
- Aydede, M. & Güzeldere, G. 2005. Cognitive architecture, concepts, and introspection: An information-theoretic solution to the problem of phenomenal consciousness. *Noûs*, 39, 197–255.
- Aydede, M. 2003. Is introspection inferential? In B. Gertler (Ed.) *Privileged access: Philosophical accounts of self-knowledge*. Burlington VT: Ashgate, 55–64.
- Baars, B. 1988. *A Cognitive Theory of consciousness*. New York: Cambridge University Press, New York.
- Baars, B. 1994. A thoroughly empirical approach to consciousness. *Psyche*, 1, <http://psyche.cs.monash.edu.au/v2/psyche-1-6-baars.html>.

- Baars, B. 1996. Understanding subjectivity: Global workspace theory and the resurrection of the observing self. *Journal of Consciousness Studies*, 3, 211–216.
- Baars, B. 1997. *In the theatre of consciousness: The workspace of the mind*. New York: Oxford University Press.
- Baars, B. 2002. The conscious access hypothesis: Origins and recent evidence. *Trends in Cognitive Sciences*, 6, 47–52.
- Baker, L. R. 1987. *Saving belief*. Princeton: Princeton University Press.
- Baker, L. R. 1995. *Explaining attitudes*. New York: Cambridge University Press.
- Baker, L. R. 1997. Why constitution is not identity? *Journal of Philosophy*, 12, 599–621.
- Baker, L. R. 1998. The first-person perspective: A test for naturalism. *American Philosophical Quarterly*, 35, 327–348.
- Baker, L. R. 2000. *Persons and bodies: A constitution view*. New York: Cambridge University Press.
- Baker, L. R. 2001. Reply to Garrett, Symposium on Persons and bodies. A field guide to philosophy of mind, http://www.uniroma3.it/kant/field/bakersymp_replytojarrett.htm.
- Baron-Cohen, S. 1995. *Mindblindness*. Cambridge, MA: MIT Press.
- Barsalou, L. W. 1987. The instability of graded structure in concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization*. Cambridge, England: Cambridge University Press, 101–140.
- Barsalou, L. W. 1993. Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In A. C. Collins, S. E. Gathercole, & M. A. Conway (Eds.) *Theories of memory*. Hillsdale, NJ: Erlbaum, 29–101.
- Barsalou, L. W. 1999. Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.
- Bechtel, W. & Mundale, J. 1999. Multiple realizability revisited: Linking cognitive and neural states. *Philosophy of science*, 66, 175–207.
- Bechtel, W. 2008. *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. London: Routledge & Kegan Paul.
- Beckerman, A. 1992. Supervenience, emergence, and reduction. In A. Beckerman, H. Flohr & J. Kim (Eds.) *Emergence or reduction: Essays on the prospects of nonreductive physicalism*. Berlin & New York: Walter de Gruyter.
- Bennett, M.R. & Hacker, P. M. S. 2003. *Philosophical foundations of neuroscience*. Oxford, England: Blackwell.
- Bermúdez, J. L. 1995a. Syntax, semantics and the levels of explanation. *Philosophical Quarterly*, 45, 361–367.
- Bermúdez, J. L. 1995b. Transcendental arguments and philosophy. *Metaphilosophy*, 26, 379–401.
- Bermúdez, J. L. 2003. The domain of folk psychology. In A. O’Hear, (Ed.) *Minds and persons*. Cambridge, England: Cambridge University Press, 25–39.

- Bermúdez, J. L. 1998. *The paradox of self-consciousness*. Cambridge, MA: MIT Press.
- Bermúdez, J. L. 2000. Personal and sub-personal: A difference without a distinction. *Philosophical Explorations*, 2, 63–82.
- Bermúdez, J. L. 2001. Abstract of “The first-person perspective: A test for naturalism”,
<http://www.philosophy.stir.ac.uk/cnw/webpapers/baker1.htm>.
- Bermúdez, J. L. 2003. *Thinking without language*. Oxford, England: Oxford University Press.
- Bermúdez, J. L. 2005. *Philosophical psychology: A contemporary introduction*. London: Routledge.
- Bermúdez, J. L. & Cahen, A. 2008. Nonconceptual mental content.
<http://plato.stanford.edu/entries/content-nonconceptual/>.
- Bermúdez, J. L., Marcel, A. & Eilan, N. (Eds.) 1997. *The body and the self*. Cambridge, MA: MIT Press.
- Blackmore, S. 1999. *The meme machine*. New York: Oxford University Press.
- Block, N. 1995. On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18, 227–47.
- Block, N. 2001. Paradoxes and cross-purposes in recent work on consciousness. *Cognition*, 79, 197–219.
- Block, N. 2002. Concepts of consciousness. In D. Chalmers (Ed.) *Philosophy of mind: Classical and contemporary Readings*. New York: Oxford University Press. pp. 206–218.
- Block, N. 2005. Two neural correlates of consciousness. *Trends in Cognitive Sciences*, 9, 46–52.
- Block, N. & Stalnaker, R. 1999. Conceptual analysis, dualism, and the explanatory gap. *The Philosophical Review*, 108, 1–46.
- Bonda, E., Petrides, M., Frey, S., & Evans, A. 1995. Neural correlates of mental transformations of the body-in-space. *Proceedings of the National Academy of Sciences*, 92, 11180–11184.
- Bowlby, J. 1969. *Attachment: Attachment and loss* (vol. 1). New York: Basic Books.
- Browne, D. 1999. Carruthers on the deficits of animals. *Psyche*, 5,
<http://psyche.cs.monash.edu.au/v5/psyche-5-23-browne.html>.
- Byrne, A. 1997. Some like it HOT: Consciousness and higher-order thoughts. *Philosophical Studies*, 86, 103–129.
- Byrne, A. 2005. Perception and conceptual content. In E. Sosa & M. Steup (Eds.) *Contemporary debates in Epistemology*. New York: Blackwell.
- Canales, A. F., Gómez, D. M. & Maffet, C. R. 2007. A critical assessment of the consciousness by synchrony hypothesis. *Biological Research*, 40, 511–519.
- Carruthers, P. 1996a. *Language, thought and consciousness: An essay in philosophical psychology*. Cambridge, England: Cambridge University Press.
- Carruthers, P. 1996b. Simulation and self-knowledge: A defense of theory-theory. In P. Carruthers & P. K. Smith (Eds.) *Theories of theories of mind*. Cambridge, England: Cambridge University Press, 22–38.

- Carruthers, P. 1997. Fragmentary versus reflexive consciousness. *Mind and Language*, 12, 181–95.
- Carruthers, P. 1998a. Natural theories of consciousness. *European Journal of Philosophy*, 6, 203–222.
- Carruthers, P. 1998b. Animal subjectivity, *Psyche*, 4, <http://psyche.cs.monash.edu.au/v4/psyche-4-03-carruthers.html>.
- Carruthers, P. 2000. *Phenomenal consciousness: a naturalistic theory*. Cambridge, England: Cambridge University Press.
- Carruthers, P. 2001. Consciousness: Explaining the phenomena. *Philosophy*, 49 (Suppl.), 61–85.
- Carruthers, P. & Smith, P. K. (Eds.) 1996. *Theories of theories of mind*. Cambridge, England: Cambridge University Press.
- Carruthers, P. 2004. HOP over FOR, HOT theory. In R. J. Gennaro (Ed.) *Higher-order theories of consciousness: An anthology*. Philadelphia: John Benjamins, 115–136.
- Castañeda, H.-N. 1999. *The Phenomeno-logic of the I*. Bloomington: Indiana University Press.
- Caston, V. 2002. Aristotle on consciousness. *Mind*, 111, 751–815.
- Chalmers, D. J. 1996. *The conscious mind*. New York: Oxford University Press.
- Chalmers, D. J. 1997. Availability: The cognitive basis of experience? *Behavioral and Brain Sciences*. 20, 148–149.
- Chalmers, D. J. 2002a. Does conceivability entail possibility? In T. Gendler, & J. Hawthorne (Eds.) *Conceivability and possibility*. New York: Oxford University Press, 145–200.
- Chalmers, D. J. 2002b. The components of content. *Philosophy of mind: Classical and contemporary readings*. New York: Oxford University Press, 608–633.
- Chalmers, D. J. 2002c. The content and epistemology of phenomenal belief. In Q. Smith & A. Jokic (Eds.) *Consciousness: New philosophical essays*. Oxford, England: Oxford University Press, 220–272.
- Chalmers, D. J. 2004. The representational character of experience. In B. Leiter (Ed.) *The Future for philosophy*. New York: Oxford University Press, 153–181.
- Chalmers, D. J. & Jackson, F. 2001. Conceptual analysis and reductive explanation. *Philosophical Review*, 110, 315–61.
- Chrisley, R. 2004. Two problems for higher-order thought theories of consciousness. ASSC-8, <http://psyche.cs.monash.edu.au/ASSC8/cc1001.html>. Abstract.
- Churchland, P. S. 1986. *Neurophilosophy: Towards a unified theory of the mind-brain*. Cambridge, MA: MIT Press.
- Churchland, P. 1995. *The engine of reason, the seat of the soul: A philosophical journey into the brain*. Cambridge, MA: MIT Press.
- Craig, A. D. 2005. Forebrain emotional asymmetry: a neuroanatomical basis? *Trends in Cognitive Sciences*, 9, 566–571.

- Craver, C. F. 2005. Beyond reduction: Mechanisms, multifield integration and the unity of neuroscience. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36, 373–395.
- Craver, C. F. 2007. *Explaining the brain*. New York: Oxford University Press.
- Crick, F. 1994. *Astonishing hypothesis: The scientific search for the soul*. London: Simon & Schuster.
- Crick, F. & Koch, C. 1990. Towards a neurobiological theory of consciousness. *Seminars in Neuroscience*, 2, 263–275.
- Damasio, A. 1989. Concepts in the brain. *Mind and Language*, 4, 24–28.
- Damasio, A. 1994. *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam Publishing.
- Damasio, A. 1999. *The feeling of what happens: Body and emotion in the making of consciousness*. London: William Heinemann.
- Davies, M. & Stone, T. (Ed.) 1995a. *Folk psychology: The theory of mind debate*. Oxford, England: Blackwell.
- Davies, M. & Stone, T. (Ed.) 1995b. *Mental simulation: Evaluations and applications*. Oxford, England: Blackwell.
- Dehaene, S., Kertszberg, M. & Changeux, J.-P. 2003. A neuronal model of a global workspace in effortful tasks. *Proceedings of the National Academy of Sciences*, 95, 14529–14534.
- Dehaene, S. & Naccache, L. 2001. Toward a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, 79, 1–37.
- Dennett, D. 1975. Conditions of personhood. In A. O. Rorty (Ed.) *The Identities of Persons*. Berkeley, CA: University of California Press, 175–196.
- Dennett, D. 1978. Towards a cognitive theory of consciousness, In D. Dennett, *Brainstorms: Philosophical essays on mind and psychology*. Cambridge, MA, Harvard University Press, 149–173.
- Dennett, D. 1991. *Consciousness explained*. Little, Brown & company.
- Dennett, D. 1993. The message is: There is no medium. *Philosophy and Phenomenological Research*, LIII, 919–931.
- Dretske, F. 1981. *Knowledge and the flow of information*. Cambridge, MA: MIT Press.
- Dretske, F. 1986. Misrepresentation. In R. Bogdan (Ed.) *Belief: Form, content, and function*. Oxford England: Oxford University Press, 17–36.
- Dretske, F. 1988. *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: MIT Press.
- Dretske, F. 1993. Conscious experience. *Mind*, 102, 263–83.
- Dretske, F. 1995. *Naturalizing the mind*. Cambridge, MA: MIT Press.
- Dretske, F. 1999. The mind's awareness of itself. *Philosophical Studies*, 95, 103–124.
- Dretske, F. 2004. Change blindness. *Philosophical Studies*, 120, 1–18.
- Edelman, G. M. 1989. *The remembered present: A biological theory of consciousness*. New York: Basic Books.
- Eliasmith, C. 2000. *How neurons mean: A neurocomputational theory of representational content*,

- http://www.arts.uwaterloo.ca/~celiasmi/courses/old_courses/Fall01/Phil600Repn/papers/eliasmith.2001.pdf. PhD Thesis.
- Evans, G. 1982. *The varieties of reference*. Oxford, England: Oxford University Press.
- Fernandez-Duque, D., Grossi, G., Thornton, I. M., Neville, H. J. 2003. Representation of change: Separate electrophysiological markers of attention, awareness, and implicit processing. *Journal of Cognitive Neuroscience*, 15, 1-17.
- Fernandez-Duque, D., & Thornton, I. M. 2000. Change detection without awareness: Do explicit reports underestimate the representation of change in the visual system? *Visual Cognition*, 7, 324-344.
- Fodor, J. 1975. *The language of thought*. Cambridge, MA: Harvard University Press.
- Fodor, J. 1983. *The modularity of mind*. Cambridge, MA: MIT Press.
- Fodor, J. 1987. *Psychosemantics*. Cambridge, MA: MIT Press.
- Fodor, J. 1998. *Concepts: Where cognitive science went wrong*. Cambridge, MA: MIT Press.
- Fodor, J. & Lepore, E. 1992. *Holism: A Shopper's Guide*. Cambridge, MA: Wiley Blackwell.
- Fonagy, P., Target, M., Gergely, G., & Jurist, E. 2002. *Affect Regulation, Mentalization and Development of the Self*. New York: Other Publications.
- Francescotti, R. 1995. Higher-order thoughts and conscious experience. *Philosophical Psychology*, 8, 239-254.
- Frank, M. The reflection model and the perception model: Two theoretical models which overlook the peculiarity of self-consciousness, and their critique by Dieter Henrich and Sydney Shoemaker. Unpublished manuscript.
- Frank, M. 1999. Self-awareness and Self-knowledge: Mental familiarity and epistemic self-ascription. *Avant Garde Critical Studies*, 23, 193-215.
- Frank, M. 2002. Self-Consciousness and self-knowledge: On some difficulties with the reduction of subjectivity. *Constellations*, 9, 390-408.
- Frankfurt, H. 1971. Freedom of the will and concept of a person. *The Journal of Philosophy*, 68, 5-20.
- Freedman, D. & Gallese, V. 2007. Motion, emotion and empathy in esthetic experience. *Trends in Cognitive Science*, 11, 197-203.
- Frege, G. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100, 25-50.
- Frith, C. D. 1992. *The cognitive neuropsychology of schizophrenia*. Hillsdale, NJ: Erlbaum.
- Frith, C. & Frith U. 1999. Interacting minds: A biological basis. *Science*, 286, 1692-1695.
- Gallagher, S. 1995. Body Schema and Intentionality. In J. L. Bermúdez, A. Marcel, & N. Eilan (Eds.) *The Body and the Self*. Cambridge, MA: MIT Press, 225-244.
- Gallagher, S. 2005. *How the body shapes the mind*. Oxford, England: Oxford University Press.

- Gallese, V. & Goldman, A. 1998. Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive science*, 12, 493–501.
- Garrett, B. 2001. The story of I: Some comments on L. R. Baker 'Persons and Bodies. A Constitution View', Symposium on 'Persons and Bodies. A Constitution View'. *A Field Guide to Philosophy of Mind* (Spring 2001), http://www.uniroma3.it/kant/field/bakersymp_garrett.htm.
- Gennaro, R. J. 1996. *Consciousness and self-consciousness: A defense of the higher-order thought theory of consciousness*. Amsterdam: John Benjamins.
- Gennaro, R. (Ed.) 2004. *Higher-order theories of consciousness: An anthology*. Philadelphia: John Benjamins.
- Gennaro, R. 2005. The HOT theory of consciousness: Between a rock and a hard place. *Journal of Consciousness Studies*, 12, 3–21.
- Gennaro, R. 2006. Between pure self-referentialism and the extrinsic HOT theory of consciousness. In U. Kriegel, & K. Williford (Eds.) 2006. *Self-representational approaches to consciousness*. Cambridge, MA: MIT Press, 221–248.
- Gertler, B. (Ed.) 2003. *Privileged access: Philosophical accounts of self-knowledge*. Burlington VT: Ashgate.
- Gibson, J. J. 1979. *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Goldman, A. 1993. Consciousness, folk psychology, and cognitive science. *Consciousness and Cognition*, 2, 365–382.
- Goldman, A. 1995. Empathy, Mind, and Morals. In M. Davies and T. Stone (Eds.), *Mental simulation: Evaluations and applications*. Oxford, England: Blackwell, 185–207.
- Gopnik, A. & Meltzoff, A. 1997. *Words, thoughts, and theories*. Cambridge, MA: MIT Press.
- Gopnik, A. 1993. How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, 16, 1–14.
- Gordon, R. 1995. Simulation without introspection or inference from me to you. In M. Davies and T. Stone (Eds.), *Mental simulation: Evaluations and applications*. Oxford, England: Blackwell, 53–67.
- Güzeldere, G. 1997. Is consciousness the perception of what passes in one's own mind? In N. Block, O. J. Flanagan, & G. Güzeldere (Eds.) *The nature of consciousness: Philosophical debates*. Cambridge, MA: MIT Press, 789–806.
- Hardcastle, V. 2004. HOT theories of consciousness: More sad tales of philosophical intuitions gone astray. In R. J. Gennaro (Ed.) *Higher-order theories of consciousness: An anthology*. Philadelphia: John Benjamins, 277–296.
- Harman, G. 1990. The intrinsic quality of experience. *Philosophical Perspectives*, 4, 31–52.
- Hayhoe, M. M., Bensinger, D. G., Ballard, D. H. 1998. Task constraints in visual working memory. *Vision Research*, 38, 125–137.

- Heal, J. 1995. How to think about thinking. In M. Davies & T. Stone. (Eds.) *Mental simulation: Evaluations and applications*. Oxford, England: Blackwell, 33–52.
- Heck, R. G. 2000. Nonconceptual content and the space of reasons. *Philosophical Review*, 109, 483–523.
- Henrich, D. 1970. Selbstbewusstsein: Kritische Einleitung in eine Theorie. In R. Bubner, K. Cramer. & R. Wiehl (Eds.) *Hermeneutik und Dialektik*. Tübingen. pp. 257–284. (English Translation 1971): *Self-Consciousness: A Critical Introduction to a Theory, Man and World*, 4, 3–28.)
- Herrmann C. S. & Knight R. T. 2001. Mechanisms of human attention: Event-related potentials and oscillations. *Neuroscience and Behavioral Reviews*, 25, 465–476.
- Horgan, T. 1993. From supervenience to superdupervenience: Meeting the demands of a material world. *Mind*, 102, 555–586.
- Huettel, S. A., Güzeldere, G., & McCarthy, G. 2001. Dissociating the neural mechanisms of visual attention in change detection using functional MRI. *Journal of Cognitive Neuroscience*, 13, 1006–1018.
- Humphrey, N. 1986. *The inner eye*. London: Faber & Faber.
- Husserl, E. 1966. *Zur Phänomenologie des inneren Zeitbewusstseins* (Husserliana, Band X). Hague: Martinus Nijhoff.
- Husserl, E. 1968. *Cartesian meditations*. Hague: Martinus Nijhoff.
- Husserl, E. 1970. *Logical investigations*. Vol I-II. London: Routledge.
- Jack, A. & Roepstorff, A. (Eds.) 2003. *Trusting the subject? The use of introspective evidence in cognitive science: Volume 1*. Exeter: Imprint Academic.
- Jack, A. & Roepstorff, A. (Eds.) 2004. *Trusting the subject? The use of introspective evidence in cognitive science: Volume 2*. Exeter: Imprint Academic.
- Jackson, F. 1982. Epiphenomenal qualia. *Philosophical Quarterly*, 32, 127–36.
- Jackson, F. 1986. What Mary did not know. *The Journal of Philosophy*, LXXXIII, 291–295.
- James, W. 1891. *Principles of psychology* vol. 1–2. New York: Dover.
- Johnston, M. 1997. Manifest kinds. *Journal of Philosophy*, 94, 564–583.
- Jurist, E. L., Slade, A. & Bergner, S. (Eds.) 2008. *Mind to mind: Infant research, neuroscience and psychoanalysis*. New York: Other Publications.
- Kant, I. 1975. *Kritik der reinen Vernunft*. Text der Ausgabe von 1781. Köln: Atlas Verlag.
- Kim, J. 1987. Strong' and 'global' supervenience revisited. *Philosophy and Phenomenological Research*, 48, 315–326.
- Kim, J. 1993. *Supervenience and mind*. New York: Oxford University Press.
- Kim, J. 1998. *Mind in a physical world*. Cambridge, MA: MIT Press.
- Koivisto, M. & Revonsuo, A. 2007. Electrophysiological correlates of visual consciousness and selective attention. *Neuroreport*, 8, 753–6.
- Kriegel, U. 2003. Consciousness as intransitive self-consciousness: Two views and an argument. *Canadian Journal of Philosophy*, 33, 103–132.

- Kriegel, U. 2004. Consciousness and self-consciousness. *The Monist*, 87, 185–209.
- Kriegel, U. 2005. Naturalizing subjective character. *Philosophy and Phenomenological Research*, 71, 23–57.
- Kriegel, U. 2006. The same-order monitoring theory of consciousness. In U. Kriegel, & K. Williford (Eds.) 2006. *Self-representational approaches to consciousness*. Cambridge, MA: MIT Press, 143–170.
- Kriegel, U. 2007. Self-consciousness. *The Internet Encyclopedia of Philosophy*, <http://www.iep.utm.edu/s/self-con.htm>.
- Kriegel, U. 2008. Real Narrow Content. *Mind and Language*, 23, 304–328.
- Kriegel, U. 2009. Self-representationalism and phenomenology. *Philosophical Studies*, 143, 357–381.
- Kriegel, U. & Williford, K. (Eds.) 2006. *Self-representational approaches to consciousness*. Cambridge, MA: MIT Press.
- Kripke, S. 1972. Naming and necessity. In D. Davidson, & G. Harman (Eds.) *Semantics of Natural Language*. Dordrecht: D. Reidel.
- Lagerspetz, O. 2002. In the industry (Review discussion of Peter Carruthers, *Phenomenal consciousness: A naturalistic theory*). *Inquiry*, 45, 541–560.
- Lamme, V. 2003. Why visual attention and awareness are different. *Trends in Cognitive Sciences*, 7, 12–18.
- Lamme, V., & Roelsetma, P. R. 2000. The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neuroscience*, 23, 571–579.
- Lau, H. C. & Passingham, R. E. 2006. Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proceedings of the National Academy of Sciences*, 103, 18763–18768.
- Laurence, S. & Margolis, E. 2006. Concepts. *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/concepts/>.
- Lehar, S. 2003. Gestalt isomorphism and the primacy of subjective conscious experience: A Gestalt Bubble model. *Behavioral and Brain Sciences*, 26, 375–444.
- Levine, J. 2001. *Purple haze*. Cambridge, MA: MIT Press.
- Lewis, C. I. 1929. *Mind and world order: Outline of a theory of knowledge*. New York: Dover.
- Llinas, R. 1996. The brain as a closed system modulated by the senses. In R. Llinàs & P. S. Churchland (Eds.) *The Mind-Brain Continuum*. Cambridge, MA: MIT Press.
- Loar, B. 1997. Phenomenal states. In N. Block, O. J. Flanagan, & G. Güzeldere (Eds.) *The nature of consciousness: Philosophical debates*. Cambridge, MA: MIT Press, 597–617.
- Locke, J. 1690. *An Essay Concerning Human Understanding*. London: William Tegg & company.
- Lurz, R. W. 2003. Neither HOT nor COLD: An alternative account of consciousness. *Psyche*, 9, <http://psyche.cs.monash.edu.au/v9/psyche-9-01-lurz.html>.

- Lurz, R. W. 2004. Either FOR or HOR: A false dichotomy. In R. J. Gennaro (Ed.) *Higher-order theories of consciousness: An anthology*. Philadelphia: John Benjamins, 227–254
- Lurz, R. W. 2002. Reducing consciousness by making it HOT: A review of Peter Carruthers' *Phenomenal consciousness*. *Psyche*, 8, http://psyche.cs.monash.edu.au/book_reviews/carruthers2/.
- Lutz, A., Lachaux, J.-P., Martinerie, J. and Varela, F. J. 2002. Guiding the study of brain dynamics by using first-person data: Synchrony patterns correlate with ongoing conscious states during a simple visual task. *Proceedings of the National Academy of Sciences*, 99, 1586–1591.
- Lycan, W. G. 1987. *Consciousness*. Cambridge, MA: MIT Press.
- Lycan, W. G. 1996. *Consciousness and experience*. Cambridge, MA: MIT Press.
- Lycan, W.G. 2001. A simple argument for a higher-order representation theory of consciousness. *Analysis*, 61, 3–4.
- Lycan, W.G. 2004. Superiority of HOP to HOT. In Gennaro, R. J. (Ed.) *Higher-order theories of consciousness: An anthology*. Philadelphia: John Benjamins, 93–104.
- Lyyra, P. 2003. *Itseymmärrys Paul Ricoeurin tekstin hermeneutiikassa*. Jyväskylä: Kampus Kustannus.
- Lyyra, P. 2004. Lynne Rudder Baker's 'first-person perspective' and the immunity to error through misidentification. In H. Ikäheimo, J. Kotkavirta, A. Laitinen, & P. Lyyra (Eds.). 2004. *Personhood: Workshop papers of the conference 'Dimensions of Personhood'*. Jyväskylä: University of Jyväskylä, Publications of Philosophy 68, 143-151.
- Lyyra, P. 2005a. Review of José Luis Bermúdez: *Thinking without words*. *Psyche* 11, http://psyche.cs.monash.edu.au/book_reviews/bemudez2/.
- Lyyra, P. 2005b. Representational emotions represented. In Gustafsson, Å. Kettunen, M., Kronqvist, C. & McEachrane, M. (Ed.) *Proceedings from the conference "Emotions, Other & Self"*. Turku: Åbo Akademi Trykkeri. (CD-ROM).
- Lyyra, P. 2009. Two senses for 'givenness of consciousness'. *Phenomenology and the Cognitive Sciences*, 8, 67–87.
- Lyyra, P., Wikgren, J. & Astikainen, P. 2010. Event-related potentials reveal rapid registration of features of infrequent changes during change blindness. *Behavioral and Brain Functions*, 6 (12), DOI: 10.1186/1744-9081-6-12.
- Mancia, M. 2007. *Feeling the words: Neuropsychanalytic understanding of memory and the unconscious*. Translated by J. Baggott. London: Routledge.
- Mandik, P. 1999. Qualia, space, and control. *Philosophical Psychology*, 12, 47 – 51.
- Marcel, A. & Bisiach, E. 1988. *Consciousness in contemporary science*. Oxford, England: Clarendon Press.
- Margolis, E. & Laurence, S. 1999. *Concepts: Core readings*. Cambridge, MA: MIT Press.

- Martin, M. 2006. On being alienated. In T. Gendler & J. Hawthorne (Eds.) *Perceptual Experience*. Oxford, England: Clarendon Press, 354–410.
- McDowell, J. 1994a. The content of perceptual experience. *Philosophical Quarterly*, 44, 190–205.
- McDowell, J. 1994b. *Mind and world: With a new introduction*. Cambridge, MA: Harvard University Press.
- Melloni, L., Molina, C., Pena, M., Torres, D., Singer, W. & Rodriguez, E. 2007. Synchronization of neural activity across cortical areas correlates with conscious perception. *Journal of Neuroscience*, 27, 2858–2865.
- Meltzoff, A. & Moore, K. 1995. Infants' understanding of people and things: From body imitation to folk psychology. In J. B. Bermudez & A. Marcel & N. Eilan (Eds.) *The body and the self*. Cambridge, MA: MIT Press, 43–69.
- Merleau-Ponty, M. 1945. *Phénoménologie de la perception*. Paris: Gallimard.
- Metzinger, T. 2004. *Being no-one*. Cambridge, MA: MIT Press.
- Milner, D. & Goodale, M. 1995. *The visual brain in action*. Oxford, England: Oxford University Press.
- Millikan, R. 1984. *Language, thought, and other biological categories: New foundations for realism*. Cambridge, MA: MIT Press.
- Nagel, T. 1974. What is it like to be a bat? *The Philosophical Review*, 82, 435–56.
- Nagel, T. 1986. *The view from nowhere*. New York: Oxford University Press.
- Neander, K. 1998. The division of phenomenal labor: A problem for representational theories of consciousness. In J. Tomberlin (Ed.), *Language, mind, and ontology*. Oxford, England: Blackwell, 411–434.
- Neisser, U. 1976. *Cognition and Reality*. San Francisco: Freeman.
- Neisser, U. 1988. Five kinds of self-knowledge. *Philosophical Psychology*, 1, 35–59.
- Nichols, S., & Stich, S. 2002. Reading one's own mind: A cognitive theory of self-awareness. In Q. Smith & A. Jokic (Eds.) *Consciousness: New philosophical essays*. New York: Oxford University Press, 157–200.
- Nisbett, R. E. & Wilson, T. D. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231–259.
- Noë, A. 2004. *Action in Perception*. Cambridge, MA: MIT Press.
- O'Brien, G. & Opie, J. 1999. A connectionist theory of phenomenal experience. *Behavioral and Brain Sciences*, 22, 127–196.
- O'Regan, K., & Noë, A. 2001. A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24, 939–1031.
- Overgaard, M. 2004. Confounding factors in contrastive analysis. *Synthese*, 141, 217–231.
- Overgaard, M. & Sørensen T. A. 2004. Introspection distinct from first-order experiences. In A. Jack, & A. Roepstorff (Eds.) *Trusting the Subject? The Use of Introspective Evidence in Cognitive Science: Volume 2*. Imprint Academic. pp. 77–95.
- Palva, S., & Palva, J. M. 2007. New vistas for alpha-frequency band oscillations. *Trends in Neuroscience*, 30, 150–158.
- Panksepp, J. 1998. *Affective neuroscience: Foundations of human and animal emotions*. New York: Oxford University Press.

- Panksepp, J. 2003. Damasio's error. *Cognition and Emotion*, 4, 111-134.
- Papineau, David 1993. *Philosophical naturalism*. Oxford, England: Blackwell.
- Papineau, David 2002. *Thinking about consciousness*. Oxford, England: Oxford University Press.
- Peacocke, Christopher 1992. *Study of concepts*. Cambridge, MA: MIT Press.
- Pepperberg, I. M. 2003. *The Alex studies: Cognitive and communicative abilities of grey parrots*. Cambridge, MA: Harvard University Press.
- Perner, Joseph 1991. *Understanding the representational mind*. Cambridge, MA: MIT Press.
- Putnam, H. 1975. The meaning of "meaning". In K. Gunderson (Ed.) *Language, Mind and Knowledge*, Vol. VII of *Minnesota Studies in the Philosophy of Science*, 131-193. Minneapolis: University of Minnesota Press.
- Pitt, D. 2004. The phenomenology of cognition, or, what is it like to think that P? *Philosophy and Phenomenological Research*, 69, 1-36.
- Prinz, J. 2000a. The duality of content. *Philosophical Studies*, 100, 1-34.
- Prinz, J. 2000b. A neurofunctional theory of visual consciousness. *Consciousness and Cognition*, 9, 243-259.
- Prinz, J. 2002. *Furnishing the mind: Concepts and their perceptual basis*. Cambridge, MA: MIT Press
- Prinz, J. 2003. A neurofunctional theory of consciousness. In A. Brook and K. Akins (Eds.). *Cognition and the brain: Philosophy and neuroscience movement*. New York: Cambridge University Press, 381-396.
- Prinz, J. 2004a. *Gut reactions: A perceptual theory of emotions*. New York: Oxford University Press.
- Prinz, J. 2004b. Fractionation of introspection. In A. Jack, & A. Roepstorff (Eds.) *Trusting the subject? The use of introspective evidence in cognitive science: Volume 2*. Exeter: Imprint Academic, 41-57.
- Prinz, J. 2007. Mental pointing. *Journal of Consciousness Studies*, 14, 184-211.
- Quine, W. V. O. 1951. Two dogmas of empiricism. *The Philosophical Review*, 60, 20-43.
- Radovic, S. 2005. *Introspecting representations*. *Acta Philosophica Gothoburgensia*, 19. Gothenburg: Acta Universitatis Gothoburgensis. PhD Thesis.
- Ramsøy, T. Z. & Overgaard M. 2004. Introspection and subliminal perception. *Phenomenology and the Cognitive Sciences*, 3, 1-23.
- Rensink, R., O'Regan, K., Clark, J. 1997. To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8, 368-373.
- Revonsuo A. 1995. *Consciousness, Dreams, and Virtual Realities*. *Philosophical Psychology*, 8, 35-58.
- Revonsuo, A. 2006. *Inner presence: Consciousness as a biological phenomenon*. Cambridge, MA: MIT Press.
- Revonsuo, A., Wilenius-Emet, M., Kuusela, J., & Lehto, M. 1997. The neural generation of a unified illusion in human vision. *Neuroreport*, 8, 3867-3870.

- Rey, G. 1983. A reason for doubting the existence of consciousness. In R. J. Davidson, G. E. Schwartz & D. Shapiro (Eds.) *Consciousness and self-regulation*. Vol. 3. New York: Plenum Press, 1–39.
- Rice, C., Koinis, D., Sullivan, K., Tager-Flusberg, H., Winner E. 1997. When 3-year olds pass the appearance-reality Test. *Developmental Psychology*, 33, 54–61.
- Ricoeur, P. 1950. *Philosophie de la volonté I. La volontaire et l'involontaire*. Paris: Aubier.
- Ricoeur, P. 1960a. *Philosophie de la volonté II. La finitude et culpabilité I. L'homme faillible*. Paris: Aubier.
- Ricoeur, P. 1960b. *Philosophie de la volonté II: La finitude et culpabilité II. La symbolique du mal*. Paris: Aubier.
- Rolls, E. 1999. *The brain and emotion*. Oxford, England: Oxford University Press.
- Rosch, E. 1975. Cognitive representation of semantic categories. *Journal of Experimental Psychology*, 104, 573–605. (Reprinted in Margolis & Laurence 1999)
- Rosenthal, D. 1986. Two concepts of consciousness. *Philosophical Studies*, 49, 329–359.
- Rosenthal, D. 1991. The independence of consciousness and sensory quality. In Villanueva, E. (Ed.) *Consciousness (Philosophical Issues, 1)*. Atascadero, CA: Ridgeview, 15–36.
- Rosenthal, D. 1993a. Multiple drafts and higher-order thoughts. *Philosophy and Phenomenological Research*, vol. LIII: 4, pp. 911–918.
- Rosenthal, D. 1993b. State and transitive consciousness. *Consciousness and Cognition*, 2, 355–363.
- Rosenthal, D. 1993c. Thinking that one thinks. In Davies, M. & Humphreys, G. (Eds.). *Consciousness: Psychological and philosophical essays*. Oxford, England: Blackwell, 197–223.
- Rosenthal, D. 1997a. Apperception, sensation, dissociability. *Mind and Language*, 12, 206–223.
- Rosenthal, D. 1997b. A theory of consciousness. In N. Block, O. J. Flanagan, & G. Güzeldere (Eds.) *The nature of consciousness: Philosophical debates*. Cambridge, MA: MIT Press, 729–754.
- Rosenthal, D. 1998a. Introspection. In R. A. Wilson, & F. C. Keil (Eds.) *MIT Encyclopedia of Cognitive Science*. Cambridge, Ma: MIT Press, 419–421.
- Rosenthal, D. 1998b. The kinds of consciousness. Delivered at the Oxford University Autumn School on Cognitive Neuroscience, October 2, 1998. <http://web.gc.cuny.edu/cogsci/kinds.pdf>
- Rosenthal, D. 2000a. Consciousness, content and metacognitive judgements. *Consciousness and Cognition*, 9, 203–214.
- Rosenthal, D. 2000b. Metacognition and Higher-Order Thoughts. *Consciousness and Cognition*, 9, 231–242.
- Rosenthal, D. 2002a. Consciousness and Higher-Order Thought. In L. Nadell (Ed.) *MacMillan Encyclopedia of Cognitive Sciences*. London: MacMillan, 717–726.

- Rosenthal, D. 2002b. Consciousness and the Mind. *Iyyun, The Jerusalem Philosophical Quarterly*, 51, 227–251.
- Rosenthal, D. 2002c. Explaining consciousness. In Chalmers, David (Ed.) *Philosophy of Mind: Classical and Contemporary Readings*. New York: Oxford University Press. 406–421.
- Rosenthal, D. 2002d. Persons, minds, and consciousness. In Auxier R. E. & Hahn L. E. (Eds.) *The Philosophy of Marjorie Grene (The Library of living philosophers)*. La Salle, Illinois: Open Court, 199–220.
- Rosenthal, D. 2003. Unity of consciousness and the self. *Proceedings of the Aristotelian Society*, 103, 325–352.
- Rosenthal, D. 2004. Varieties of higher-order thought. In R. J. Gennaro (Ed.) *Higher-order theories of consciousness: An anthology*. Philadelphia: John Benjamins, 67–92., 17–44.
- Rosenthal, D. 2005. *Consciousness and mind*. New York: Oxford University Press.
- Rosenthal, D. 2006. Tutorial on Higher-Order Theories of Consciousness, Delivered at Oxford, the annual meeting of the Association for the Scientific Study of Consciousness (ASSC) June 10, 2006.
<http://davidrosenthal1.googlepages.com/assc-10.pdf>.
- Rosenthal, D. 2008. Consciousness and its function. *Neuropsychologia*, 829–840.
- Saariluoma, P. 1997. *Foundational analysis*. London: Routledge.
- Sapp, F., Lee, K. & Muir, D. 2000. Three-year-olds' difficulty with the appearance–reality distinction: Is it real or is it apparent? *Developmental Psychology*, 36, 547–560.
- Schröder, J. 2001. Higher order thought and naturalist accounts of consciousness. *Journal of Consciousness Studies*, 8, 27–46.
- Schroeder, T. 2004. *Three faces of desire*. New York: Oxford University Press.
- Seager, W. 1994. Dretske on HOT theories of consciousness. *Analysis*, 54, 207–276.
- Seager, W. 2002. Emotional Introspection. *Consciousness and Cognition*, 11, 666–687.
- Seager, W. & Bourget, D. 2007. Representationalism about consciousness. In M. Velmans and S. Schneider (Eds.) *Blackwell Companion to Consciousness*, Oxford, England: Blackwell, 261–276.
- Searle, J. 1983. *Intentionality: An essay in the philosophy of mind*. New York: Cambridge University Press.
- Searle, J. 1992. *Rediscovery of mind*. Cambridge, MA: MIT Press.
- Sellars, W. 1956. Empiricism and the philosophy of mind. In H. Feigl & M. Scriven (Eds.) *Minnesota Studies in the Philosophy of Science, Volume I: The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, University of Minnesota Press, 253–329.
- Shoemaker, S. 1994. Self-knowledge and "inner sense", *Philosophy and Phenomenological Research*, LIV, 249–314.
- Shoemaker, S. 1994b. Self-reference and self-awareness, In Cassam, Q. (Ed.) *Self-knowledge*. Oxford, England: Oxford University Press, 80–93.

- Shoemaker, S. 1996. *The first-person perspective and other essays*. Cambridge, England: Cambridge University Press.
- Shoemaker, S. 1997. Phenomenal character. In A. Byrne & D. R. Hilbert (Eds.) *Readings on color: The philosophy of color*, Cambridge, MA: MIT Press, 227-245.
- Simons, D. J., & Levin, D. T. 1997. Change Blindness. *Trends in Cognitive Science*, 1, 261-267.
- Smith, B. Husserlian ecology. Unpublished manuscript, <http://ontology.buffalo.edu/smith/articles/husserlianecology.html>.
- Smith, J. D., Shields, W. E., Washburn, D. A. 2003. The comparative psychology of uncertainty monitoring and metacognition. *Behavioral and Brain Sciences*, 26, 317-373.
- Smith, D. W. 1989 *Circle of acquaintance*. Dordrecht: Kluwer Academic.
- Smith, D. W. 2003. Return to consciousness. In D. W. Smith (Ed.) *Mind and world: Essays in phenomenology and ontology*. Cambridge, England: Cambridge University Press.
- Spelke, E. S. 1990. Principles of object perception. *Cognitive Science*, 14, 29-56.
- Stoerig, P. 2005. Consciousness as state and trait. Delivered at TSC-2005, University of Copenhagen, August, 19.
- Strawson, G. 1994. *Mental reality*. Cambridge, MA: MIT Press.
- Strawson, P. 1959. *Individuals: An essay in descriptive metaphysics*. London: Methuen.
- Stubenberg, L. 1998. *Consciousness and qualia*. Amsterdam: John Benjamins.
- Thagard, P. 2008. *Cognitive science. Routledge companion to philosophy of science*. London: Routledge, 531-542.
- Thomas, A. 2003. An adverbial theory of consciousness. *Phenomenology and the Cognitive Sciences*, 2, 161-185.
- Thomasson, A. 2000. After Brentano: A one-level theory of consciousness. *European Journal of Philosophy*, 8, 190-210.
- Thomasson, A. 2005. First-person knowledge in phenomenology. In D. W. Smith, & A. Thomasson (Eds.) *Phenomenology and philosophy of mind*. Oxford, England: Oxford University Press, 115-138.
- Thomasson, A. 2006. Self-awareness and self-knowledge. *Psyche*, 12, <http://psyche.cs.monash.edu.au/>.
- Thomasson, A. 2008. Phenomenal consciousness and phenomenal world. *Monist*, 91, 191-214.
- Titchener, E. 1912a. Prolegomena to a study of introspection. *The American Journal of Psychology*, 23, 427-448.
- Titchener, E. 1912b. The schema of introspection. *The American Journal of Psychology*, 23, 485-508.
- Tsakiris, M., Hesse, M. D., Boy, C., Haggard, P. & Fink, G. R. 2007. Neural signatures of body ownership: a sensory network for bodily self-consciousness. *Cerebral Cortex*, 17, 2235-2244.
- Tye, M. 1995. *Ten problems of consciousness: A representational theory of the phenomenal mind*. Cambridge, MA: MIT Press.
- Tye, M. 2000. *Consciousness, color and content*. Cambridge, MA: MIT Press.

- Tye, M. 2002. Representationalism and the transparency of experience. *Noûs*, 36, 137–151.
- Van Gulick, R. 2000. Inward and upward: Reflection, introspection, and self-awareness, *Philosophical Topics*, 28, 275–305.
- Van Gulick, R. 2003. Maps, gaps and traps. In A. Jokic and Q. Smith (Eds). *Consciousness: New philosophical perspectives*. Oxford, England: Oxford University Press.
- Van Gulick R. 2004a. Consciousness. *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/consciousness/>.
- Van Gulick, R. 2004b. Higher-order global states: An alternative model of higher order model of consciousness. In R. J. Gennaro (Ed.) *Higher-order theories of consciousness: An anthology*. Philadelphia: John Benjamins, 67–92.
- Varela, F. 1996. Neurophenomenology : A Methodological Remedy for the Hard Problem. *Journal of Consciousness Studies*, 3 , 330-349.
- Varley, R. 1998. Aphasic language, aphasic thought. In Carruthers, P. and Boucher, J. (Eds.) *Language and Thought*. Cambridge, England: Cambridge University Press, 128–145.
- Varley, R., Siegal, M., Want, S. C. 2001. Severe impairment in grammar does not preclude theory of mind. *Neurocase*, 7, 489–93.
- Vermersch, P. 1999. Introspection as practice. *Journal of Consciousness Studies*, 6, 17–42.
- Vogele, K., Bussfeld, P., Newen, A., Herrmann, S., Happe, F., Falkai, P., Maier, W., Shah, N. J., Fink, G. R. & Zilles, K. 2001. Mind reading: Neural mechanisms of theory of mind and self-perspective, *NeuroImage*, 14, 170 – 181.
- Weisberg, J. 1999. Active, thin and hot! An actualist response to Carruthers' dispositionalist HOT view. *Psyche*, 5, <http://psyche.cs.monash.edu.au/v5/psyche-5-06-weisberg.html>.
- Weiskrantz, L. 1986. *Blindsight: A case study and implications*. Oxford, England: Clarendon Press.
- Weiskrantz, L. 1997. *Consciousness lost and found*. Oxford, England: Oxford University Press, Oxford and New York.
- Wiggins, D. 2001. *Sameness and substance renewed*. Cambridge, England: Cambridge University Press.
- Wilkes, K. 1988. *Real people: Personal identity without thought experiments*. Oxford, England: Clarendon University Press.
- Williford, K. 2006. Zahavi versus Brentano: A rejoinder. *Psyche*, 12, <http://psyche.cs.monash.edu.au/symposia/kriegel/R4Williford.pdf>.
- Wimmer, H. & Perner, J. 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103-128.
- Wittgenstein, L. 1953. *Philosophical investigations*. Oxford, England: Basil Blackwell.
- Wittgenstein, L. 1974. *The blue and brown books: Preliminary studies for the 'Philosophical investigations'*. Oxford, England: Wiley Blackwell.

- Woodward, J. 2003. *Making things happen: A theory of causal explanation*. Oxford, England: Oxford University Press.
- Wright, W. 2003. McDowell, demonstrative concepts, and nonconceptual representational content. *Disputatio*, 14, 37–51.
- Yrjönsuuri, M. 2007. The structure of self-consciousness: A fourteenth-century debate. In S. Heinämaa, V. Lähteenmäki & P. Remes (Eds.) *Consciousness: From Perception to Reflection in the History of Philosophy (Studies in the History of Philosophy of Mind [4])*. Dordrecht: Springer, 141–152.
- Zahavi, D. 1998. *Self-Awareness and alterity: a phenomenological investigation*. Illinois: Northwestern University Press.
- Zahavi, D. (Ed.) 2000. *Exploring the self. Advances in Consciousness Research 23*. Amsterdam-Philadelphia: John Benjamins.
- Zahavi, D. 2003. *Husserl's phenomenology*. Stanford: Stanford University Press.
- Zahavi, D. 2004. Husserl's noema and the internalism–externalism debate. *Inquiry*, 47, 42–66.
- Zahavi, D. 2005. *Subjectivity and selfhood*. Cambridge, MA: MIT Press.
- Zahavi, D. & Parnas, J. 1999. Phenomenal Consciousness and Self-Awareness: A Phenomenological Critique of Representational Theory. In S. Gallagher & J. Shear (Eds.) *Models of the Self*. Exeter: Imprint Academic, 253–270.
- Zeki, S. 2003. The disunity of consciousness. *Trends in Cognitive Sciences*, 7, 214–218.

JYVÄSKYLÄ STUDIES IN EDUCATION, PSYCHOLOGY AND SOCIAL RESEARCH

- 1 KYÖSTIÖ, O. K., Oppilaiden terveydentilan riippuvuus koulutyypistä. - Health status of pupils according to type of school. 78 p. Summary 6 p. 1962.
- 2 HEINONEN, VEIKKO, Differentiation of primary mental abilities. 136 p. 1963.
- 3 ELONEN, ANNA S., TAKALA, MARTTI & RUOPPILA ISTO, A study of intellectual functions in children by means of the KTK performance scales. 181 p. 1963.
- 4 JUURMAA, JYRKI, On the ability structure of the deaf. 114 p. 1963.
- 5 HEINONEN, VEIKKO, Lyhennetty faktori-analyysi. - A short method for factor analysis. 76 p. Summary 5 p. 1963.
- 6 PITKÄNEN, PENTTI, Fyysisen kunnon rakenne ja kehittyminen. - The structure and development of physical fitness. 163 p. Summary 10 p. 1964.
- 7 NURMI, VELI, Maamme seminaarien varsinaisen opettajakoulutuksen synty ja kehittyminen viime vuosisadalla I. - Die Entehung und Entwicklung der praktischen Lehrer-bildung unserer Lehrerseminare im vorigen Jahrhundert I. 270 p. Zusammenfassung 15 p. 1964.
- 8 NURMI, VELI, Maamme seminaarien varsinaisen opettajakoulutuksen synty ja kehittyminen viime vuosisadalla II. - Die Entstehung und Entwicklung der praktischen Lehrer-bildung unserer Lehrerseminare im vorigen Jahrhundert II. 123 p. Zusammenfassung 10 p. 1964.
- 9 NUMMENMAA, TAPIO, The language of the face. 66 p. 1964.
- 10 ISOSAARI, JUSSI, Bruno Boxström ja Sortavalan seminaarin kasvatusaineiden opetus 1882-1917. - Bruno Boxström und der Unterricht in den pädagogischen Fächern am Seminar von Sortavala 1882-1917. 167 p. Zusammenfassung II p. 1964.
- 11 NUMMENMAA, TAPIO & TAKALA, MARTTI, Parental behavior and sources of information in different social groups. 53 p. 1965.
- 12 WECKROTH, JOHAN, Studies in brain pathology and human performance I. - On the relationship between severity of brain injury and the level and structure of intellectual performance. 105 p. 1965.
- 13 PITKÄNEN, PENTTI, Ärsyke- ja reaktioanalyttisten faktorointitulosten vastaavuudesta. - On the congruence and coincidence between stimulus analytical and response analytical factor results. 223 p. Summary 14 p. 1967.
- 14 TENKKU, JUSSI, Are single moral rules absolute in Kant's ethics? 31 p. 1967.
- 15 RUOPPILA, ISTO, Nuorten ja varttuneiden opiskelijoiden väliset asenne-erot eräissä ylioppilaspohjaisissa oppilaitoksissa. - Attitude differences between young and advanced university and college students. 182 p. Summary 14 p. 1967.
- 16 KARVONEN, JUHANI, The structure, arousal and change of the attitudes of teacher education students. 118 p. 1967.
- 17 ELONEN, ANNA S., Performance scale patterns in various diagnostic groups. 53 p. 1968.
- 18 TUOMOLA, UUNO, Kansakouluntarkastajaan kohdistuvista rooli-odotuksista. - On role-expectations applied to school inspectors. 173 p. Summary 8 p. 1968.
- 19 PITKÄNEN, LEA, A descriptive model of aggression and nonaggression with applications to childrens behaviour. 208 p. 1969.
- 20 KOSKIAHO, BRIITTA, Level of living and industrialisation. 102 p. 1970.
- 21 KUUSINEN, JORMA, The meaning of another person's personality. 28 p. 1970.
- 22 VIJANEN, ERKKI, Pohjakoulutustaso ja kansakoulunopettajan kehitysympäristöjen muodostuminen. - The level of basic education in relation to the formation of the development milieus of primary school teachers. 280 s. Summary 13 p. 1970.
- 23 HAGFORS, CARL, The galvanic skin response and its application to the group registration of psychophysiological processes. 128 p. 1970.
- 24 KARVONEN, JUHANI, The enrichment of vocabulary and the basic skills of verbal communication. 47 p. 1971.
- 25 SEPPÖ, SIMO, Abiturienttien asenteet uskonnon-opetukseen. - The attitudes of students toward religious education in secondary school. 137 p. Summary 5 p. 1971.
- 26 RENKO MANU, Opettajan tehokkuus oppilaiden koulusaavutusten ja persoonallisuuden kehittämiseksi. - Teacher's effectiveness in improving pupils' school achievements and developing their personality. 144 p. Summary 4 p. 1971.
- 27 VAHERVA, TAPIO, Koulutustulokset peruskoulun ala-asteella yhteisömuuttajien selittäminä. - Educational outcomes at the lower level of the comprehensive school in the light of ecological variables. 158 p. Summary 3 p. 1974.
- 28 OLKINUORA, ERKKI, Norm socialization. The formation of personal norms. 186 p. Tiivistelmä 4 p. 1974.
- 29 LIIKANEN, PIIRKKO, Increasing creativity through art education among pre-school children. 44 p. Tiivistelmä 4 p. 1975.
- 30 ELONEN, ANNA S., & GUYER, MELVIN, Comparison of qualitative characteristics of human figure drawings of Finnish children in various diagnostic categories. 46 p. Tiivistelmä 3 p. 1975.
- 31 KÄÄRIÄINEN, RISTO, Physical, intellectual, and personal characteristics of Down's syndrome. 114 p. Tiivistelmä 4 p. 1975.
- 32 MÄÄTTÄ, PAULA, Images of a young drug user. 112 p. Tiivistelmä 11 p. 1976.
- 33 ALANEN, PENTTI, Tieto ja demokratia. - Epistemology and democracy. 140 p. Summary 4 p. 1976.
- 34 NUPPONEN, RIITTA, Vahvistajaroolit aikuisten ja lapsen vuorovaikutuksessa. - The experimental roles of reinforcing agent in adult-child interaction. 209 p. Summary 11 p. 1977.

- 35 TEIKARI, VEIKKO, Vigilanssi-ilmiön mittaamisesta ja selitysmahdollisuuksista. – On measuring and explanation of vigilance. 163 p. Summary 2 p. 1977.
- 36 VOLANEN, RISTO, On conditions of decision making. A study of the conceptual foundations of administration. – Päätöksenteon edellytyksistä. Tutkimus hallinnon käsitteellisistä perusteista. 171 p. Tiivistelmä 7 p. 1977.
- 37 LYYTINEN, PAULA, The acquisition of Finnish morphology in early childhood. – Suomen kielen morfologisten säännönmukaisuuksien omaksuminen varhaislapsuudessa. 143 p. Tiivistelmä 6 p. 1978.
- 38 HAKAMÄKI, SIMO, Maaseudulle muutto muutto liikkeen osana. – Migration on rural areas as one element of migration as a whole. 175 p. Summary 5 p. 1978.
- 39 MOBERG, SAKARI, Leimautuminen erityispedagogiikassa. Nimikkeisiin apukoululainen ja tarkkailuluokkainen liittyvät käsitykset ja niiden vaikutus hypoteettista oppilasta koskeviin havaintoihin. – Labelling in special education. 177 p. Summary 10 p. 1979.
- 40 AHVENAINEN, OSSI, Lukemis- ja kirjoittamishäiriöinen erityisopetuksessa. – The child with reading and writing disabilities in special education. 246 p. Summary 14 p. 1980.
- 41 HURME, HELENA, Life changes during childhood. – Lasten elämänmuutokset. 229 p. Tiivistelmä 3 p. 1981.
- 42 TUTKIMUS YHTEISKUNTAPOLITIIKAN VIITOITAJANA. Professori Leo Paukkuselle omistettu juhla kirja. 175 p. 1981.
- 43 HIRSJÄRVI, SIRKKA, Aspects of consciousness in child rearing. – Tietoisuuden ongelma kotikasvatuksessa. 259 p. 1981.
- 44 LASONEN, KARI, Siirtolaisoppilas Ruotsin kouluyhteisössä. Sosiometrinen tutkimus. – A socio-metric study of immigrant pupils in the Swedish comprehensive school. 269 p. Summary 7 p. 1981.
- 45 AJATUKSEN JA TOIMINNAN TIET. Matti Juntusen muistokirja. 274 p. 1982.
- 46 MÄKINEN, RAIMO, Teachers' work, wellbeing, and health. – Opettajan työ, hyvinvointi ja terveys. 232 p. Tiivistelmä 2 p. 1982.
- 47 KANKAINEN, MIKKO, Suomalaisen peruskoulun eriyttämisratkaisun yhteiskunnallisen taustan ja siirtymävaiheen toteutuksen arviointi. 257 p. Summary 11 p. 1982.
- 48 WALLS, GEORG, Health care and social welfare in, cooperation. 99 p. Tiivistelmä 9 p. 1982.
- 49 KOIVUKARI, MIRJAMI, Rote learning comprehension and participation by the learners in Zairian classrooms. – Mekaaninen oppiminen, ymmärtäminen ja oppilaiden osallistumisen opetukseen zairelaisissa koululuokissa. 286 p. Tiivistelmä 11 p. 1982.
- 50 KOPONEN, RITVA, An item analysis of tests in mathematics applying logistic test models. – Matematiikan kokeiden osioanalyysi logistisia testimalleja käyttäen. 187 p. Tiivistelmä 2 p. 1983.
- 51 PEKONEN, KYÖSTI, Byrokrania politiikan näkökulmasta. Poliitiikan ja byrokratian keskinäinen yhteys valtio- ja yhteiskuntaprosessin kehityksen valossa. – Bureaucracy from the viewpoint of politics. 253 p. 1983.
- 52 LYYTINEN, HEIKKI, Psychophysiology of anticipation and arousal. – Antisipaation ja viriämisen psykofysiologia. 190 p. Tiivistelmä 4 p. 1984.
- 53 KORAKIANGAS, MIKKO, Lastenneuvolan terveydenhoitajan arvioinnit viisivuotiaiden lasten psyykkisestä kehityksestä. – The psychological assessment of five-year-old children by public health centres. 227 p. Summary 14 p. 1984.
- 54 HUMAN ACTION AND PERSONALITY. Essays in honour of Martti Takala. 272 p. 1984.
- 55 MATILAINEN, JOUKO, Maanpuolustus ja eduskunta. Eduskuntaryhmien kannanotot ja koheesio maanpuolustuskysymyksissä Paasikiven-Kekkonen kaudella 1945-1978. – Defence and Parliament. 264 p. Summary 7 p. 1984.
- 56 PUOLUE, VALTIO JA EDUSTUKSELLINEN DEMOKRATIA. Pekka Nyholmille omistettu juhla kirja. – Party, state and representational democracy. 145 p. Summary 2 p. 1986.
- 57 SIISIÄINEN, MARTTI, Intressit, yhdistyslaitos ja poliittisen järjestelmän vakaisuus. – Interests, voluntary associations and the stability of the political system. 367 p. Summary 6 p. 1986.
- 58 MATTLAR, CARL-ERIK, Finnish Rorschach responses in cross-cultural context: A normative study. 166 p. Tiivistelmä 2 p. 1986.
- 59 ÄYSTÖ, SEIJA, Neuropsychological aspects of simultaneous and successive cognitive processes. – Rinnakkaisen ja peräkkäisen informaation prosessoinnin neuropsykologiasta. 205 p. Tiivistelmä 10 p. 1987.
- 60 LINDH, RAIMO, Suggestiiviset mielikuvamallit käyttäytymisen muokkaajina tarkkailuluokkalaisilla. – Suggestive covert modeling as a method with disturbed pupils. 194 p. Summary 8 p. 1987.
- 61 KORHONEN, TAPANI, Behavioral and neural short-latency and long-latency conditioned responses in the cat. – Välittömät ja viivästetyt hermostolliset ja käyttäytymisvasteet klassisen ehdollistamisen aikana kissalla. 198 p. Tiivistelmä 4 p. 1987.
- 62 PAHKINEN, TUULA, Psykoterapian vaikutus minäkäsitykseen. Psykoterapian käynnistämisen muutosprosessin vaikutus korkeakouluopiskelijoiden minäkäsitykseen. – Change in self-concept as a result of psychotherapy. 172 p. Summary 6 p. 1987.
- 63 KANGAS, ANITA, Keski-Suomen kulttuuri-toimintakokeilu tutkimuksena ja politiikkana. – The action research on cultural-activities in the Province of Central Finland. 301 p. Summary 8 p. 1988.
- 64 HURME, HELENA, Child, mother and grandmother. Intergenerational interaction in

- 65 RASKU-PUTTONEN, HELENA, Communication between parents and children in experimental situations. - Vanhempien ja lasten kommunikointi strukturoiduissa tilanteissa. 71 p. Tiivistelmä 5 p. 1988.
- 66 TOSKALA, ANTERO, Kahvikuppineurootikkojen ja paniikkiagorafobikkojen minäkuvat minäsystemin rakenteina ja kognitiivisen oppimis-terapian perustana. - The self-images of coffee cup neurotics and panic agoraphobics as structures of a selfsystem and a basis for learning therapy. 261 p. Summary 6 p. 1988.
- 67 HAKKARAINEN, LIISA, Kuurojen yläasteen oppilaiden kirjoitetun kielen hallinta. - Mastery of written language by deaf pupils at the upper level of Comprehensive school. 281 p. Summary 11 p. 1988.
- 68 NÄTTI, JOUKO, Työmarkkinoiden lohkoutuminen. Segmentaatioteoriat, Suomen työmarkkinat ja yritysten työvoimastrategiat. - Segmentation theories, Finnish labour markets and the use of labour in retail trade. 189 p. Summary 10 p. 1989.
- 69 AALTOLA, JUHANI, Merkitys opettamisen ja oppimisen näkökulmasta Wittgensteinin myöhäisfilosofian ja pragmatismien valossa. - Meaning from the point of view of teaching and learning in the light of Wittgenstein's later philosophy and pragmatism. 249 p. Summary 6 p. 1989.
- 70 KINNUNEN, ULLA, Teacher stress over a school year. - Opettajan työstressi lukuvuoden aikana. 61 p. Tiivistelmä 3 p. 1989.
- 71 BREUER, HELMUT & RUOHO, KARI (Hrsg.), Pädagogisch-psychologische Prophylaxe bei 4-8-jährigen Kindern. - Pedagogis-psykologinen ennaltaehkäisy neljästä kahdeksaan vuoden iässä. 185 S. Tiivistelmä 1 S. 1989.
- 72 LUMMELAHTI, LEENA, Kuusivuotiaiden sopeutuminen päiväkotiin. Yksilöllistetty mallioppimis-ohjelma päiväkotiin heikosti sopeutuvien kuusivuotiaiden ohjauksessa sekä vanhempien kasvatuskäytännön yhteydet lapsen sopeutumiseen ja minäkäsitykseen. - The adjustment of six-year-old children to day-care-centres. 224 p. Summary 9 p. 1990.
- 73 SALOVIITA, TIMO, Adaptive behaviour of institutionalized mentally retarded persons. - Laitoksessa asuvien kehitysvammaisten adaptiivinen käyttäytyminen. 167 p. Tiivistelmä 4 p. 1990.
- 74 PALONEN, KARI et SUBRA, LEENA (Eds.), Jean-Paul Sartre - un philosophe du politique. - Jean-Paul Sartre - poliittisuuden filosofi. 107 p. Tiivistelmä 2 p. 1990.
- 75 SINIVUO, JUHANI, Kuormitus ja voimavarat upseerin uralla. - Work load and resources in the career of officers. 373 p. Summary 4 p. 1990.
- 76 PÖLKKI, PIRJO, Self-concept and social skills of school beginners. Summary and discussion. - Koulutulokkaiden minäkäsitys ja sosiaaliset taidot. 100 p. Tiivistelmä 6 p. 1990.
- 77 HUTTUNEN, JOUKO, Isän merkitys pojan sosiaaliselle sukupuolelle. - Father's impact on son's gender role identity. 246 p. Summary 9 p. 1990.
- 78 AHONEN, TIMO, Lasten motoriset koordinaatiohäiriöt. Kehitysneuro psykologinen seuranta-tutkimus. - Developmental coordination disorders in children. A developmental neuropsychological follow-up study. 188 p. Summary 9 p. 1990.
- 79 MURTO, KARI, Towards the well functioning community. The development of Anton Makarenko and Maxwell Jones' communities. - Kohti toimivaa yhteisöä. Anton Makarenkon ja Maxwell Jonesin yhteisöjen kehitys. 270 p. Tiivistelmä 5 p. Cp2`<, 5 c. 1991.
- 80 SEIKKULA, JAAKKO, Perheen ja sairaalan rajasynteesi potilaan sosiaalisessa verkostossa. - The family-hospital boundary system in the social network. 285 p. Summary 6 p. 1991.
- 81 ALANEN, ILKKA, Miten teoretisoida maa-talouden pientuotantoa. - On the conceptualization of petty production in agriculture. 360 p. Summary 9 p. 1991.
- 82 NIEMELÄ, EINO, Harjaantumisoppilas peruskoulun liikuntakasvatuksessa. - The trainable mentally retarded pupil in comprehensive school physical education. 210 p. Summary 7 p. 1991.
- 83 KARILA, IRMA, Lapsivuodeajan psyykkisten vaikeuksien ennakointi. Kognitiivinen malli. - Prediction of mental distress during puerperium. A cognitive model. 248 p. Summary 8 p. 1991.
- 84 HAAPASALO, JAANA, Psychopathy as a descriptive construct of personality among offenders. - Psykopatia rikoksentekijöiden persoonallisuutta kuvaavana konstruktiona. 73 p. Tiivistelmä 3 p. 1992.
- 85 ARNKIL, ERIK, Sosiaalityön rajasynteemit ja kehitysvyöhyke. - The systems of boundary and the developmental zone of social work. 65 p. Summary 4 p. 1992.
- 86 NIKKI, MAIJA-LIISA, Suomalaisen koulutusjärjestelmän kielikoulutus ja sen relevanssi. Osa II. - Foreign language education in the Finnish educational system and its relevance. Part 2. 204 p. Summary 5 p. 1992.
- 87 NIKKI, MAIJA-LIISA, The implementation of the Finnish national plan for foreign language teaching. - Valtakunnallisen kielenopetuksen yleissuunnitelman toimeenpano. 52 p. Yhteenveto 2 p. 1992.
- 88 VASKILAMPI, TUULA, Vaihtoehtoinen terveydenhuolto hyvinvointivaltion terveystarkkailuilla. - Alternative medicine on the health market of welfare state. 120 p. Summary 8 p. 1992.
- 89 LAAKSO, KIRSTI, Kouluvaikeuksien ennustaminen. Käyttäytymishäiriöt ja kielelliset vaikeudet peruskoulun alku- ja päättövaiheissa. -

- Prediction of difficulties in school. 145 p. Summary 4 p. 1992.
- 90 SUUTARINEN, SAKARI, Herbartilainen pedagoginen uudistus Suomen kansakoulussa vuosisadan alussa (1900-1935). - Die Herbart'sche pädagogische Reform in den finnischen Volksschulen zu Beginn dieses Jahrhunderts (1900-1935). 273 p. Zusammenfassung 5 S. 1992.
- 91 AITTOLA, TAPIO, Uuden opiskelijatyypin synty. Opiskelijoiden elämänvaiheet ja tieteenalaspesifien habitusten muovautuminen 1980-luvun yliopistossa. - Origins of the new student type. 162 p. Summary 4 p. 1992
- 92 KORHONEN, PEKKA, The origin of the idea of the Pacific free trade area. - Tyynenmeren vapaa-kauppa-alueen idean muotoutuminen. - Taiheiyo jiyuu boeeki chi-iki koosoo no seisei. 220 p. Yhteenveto 3 p. Yooyaku 2 p. 1992.
- 93 KERÄNEN, JYRKI, Avohoitoon ja sairaalahoitoon valikoituminen perhekeskeisessä psykiatrisessa hoitojärjestelmässä. - The choice between outpatient and inpatient treatment in a family centred psychiatric treatment system. 194 p. Summary 6 p. 1992.
- 94 WAHLSTRÖM, JARL, Merkitysten muodostuminen ja muuttuminen perheterapeuttisessa keskustelussa. Diskurssianalyttinen tutkimus. - Semantic change in family therapy. 195 p. Summary 5 p. 1992.
- 95 RAHEEM, KOLAWOLE, Problems of social security and development in a developing country. A study of the indigenous systems and the colonial influence on the conventional schemes in Nigeria. - Sosiaaliturvan ja kehityksen ongelmia kehitysmaassa. 272 p. Yhteenveto 3 p. 1993.
- 96 LAINE, TIMO, Aistisuus, kehollisuus ja dialogisuus. Ludwig Feuerbachin filosofian lähtökohtia ja niiden kehitysnäkymiä 1900-luvun antropologisesti suuntautuneessa fenomenologiassa. - Sensuousness, bodiliness and dialogue. Basic principles in Ludwig Feuerbach's philosophy and their development in the anthropologically oriented phenomenology of the 1900's. 151 p. Zusammenfassung 5 S. 1993.
- 97 PENTTONEN, MARKKU, Classically conditioned lateralized head movements and bilaterally recorded cingulate cortex responses in cats. - Klassisesti ehdollistetut sivuttaiset päänliikkeet ja molemminpuoliset aivojen pihtipoimun vasteet kissalla. 74 p. Yhteenveto 3 p. 1993.
- 98 KORO, JUKKA, Aikuinen oman oppimisensa ohjaajana. Itseohjautuvuus, sen kehittyminen ja yhteys opetustuloksiin kasvatustieteen avoimen korkeakouluopetuksen monimuotokokeilussa. - Adults as managers of their own learning. Self-directiveness, its development and connection with the cognitive learning results of an experiment on distance education for the teaching of educational science. 238 p. Summary 7 p. 1993.
- 99 LAIHIALA-KANKAINEN, SIRKKA, Formaalin ja funktionaalinen traditio kieltenopetuksessa. Kieltenopetuksen oppihistoriallinen tausta antiikista valistukseen. - Formal and functional traditions in language teaching. The theory -historical background of language teaching from the classical period to the age of reason. 288 p. Summary 6 p. 1993.
- 100 MÄKINEN, TERTTU, Yksilön varhaiskehitys koulunkäynnin perustana. - Early development as a foundation for school achievement. 273 p. Summary 16 p. 1993.
- 101 KOTKAVIRTA, JUSSI, Practical philosophy and modernity. A study on the formation of Hegel's thought. - Käytännöllinen filosofia ja modernisuus. Tutkielma Hegelin ajattelun muotoutumisesta. 238 p. Zusammenfassung 3 S. Yhteenveto 3 p. 1993.
- 102 EISENHARDT, PETER L., PALONEN, KARI, SUBRA, LEENA, ZIMMERMANN RAINER E. (Eds.), Modern concepts of existentialism. Essays on Sartrean problems in philosophy, political theory and aesthetics. 168 p. Tiivistelmä 2 p. 1993.
- 103 KERÄNEN, MARJA, Modern political science and gender. A debate between the deaf and the mute. - Moderni valtio-oppi ja nainen. Mykkien ja kuuroiden välinen keskustelu. 252 p. Tiivistelmä 4 p. 1993.
- 104 MATIKAINEN, TUULA, Työtaitojen kehittyminen erityisammattikouluvaiheen aikana. - Development of working skills in special vocational school. 205 p. Summary 4 p. 1994.
- 105 PIHLAJARINNE, MARJA-LEENA, Nuoren sairastuminen skitsofreeniseen häiriöön. Perheterapeuttinen tarkastelutapa. - The onset of schizophrenic disorder at young age. Family therapeutic study. 174 p. Summary 5 p. 1994.
- 106 KUUSINEN, KIRSTI-LIISA, Psykkinen itsesäätely itsehoidon perustana. Itsehoito I-tyypin diabetesta sairastavilla aikuisilla. - Self-care based on self-regulation. Self-care in adult type I diabetics. 260 p. Summary 17 p. 1994.
- 107 MENGISTU, LEGESSE GEBRESELLASSIE, Psychological classification of students with and without handicaps. A tests of Holland's theory in Ethiopia. 209 p. 1994.
- 108 LESKINEN, MARKKU (ED.), Family in focus. New perspectives on early childhood special education. 158 p. 1994.
- 109 LESKINEN, MARKKU, Parents' causal attributions and adjustment to their child's disability. - Vanhempien syytulkinnat ja sopeutuminen lapsensa vammaisuuteen. 104 p. Tiivistelmä 1 p. 1994.
- 110 MATTHIES, AILA-LEENA, Epävirallisen sektorin ja hyvinvointivaltion suhteiden modernisoituminen. - The informal sector and the welfare state. Contemporary relationships. 63 p. Summary 12 p. 1994.
- 111 AITTOLA, HELENA, Tutkimustyön ohjaus ja ohjaussuhteet tieteellisessä jatkokoulutuksessa. - Mentoring in postgraduate education. 285 p. Summary 5 p. 1995.
- 112 LINDÉN, MIRJA, Muuttuva syövän kuva ja kokeminen. Potilaiden ja ammattilaisten tulkintoja. - The changing image and experience

- of cancer. Accounts given by patients and professionals. 234 p. Summary 5 p. 1995.
- 113 VÄLIMAA, JUSSI, Higher education cultural approach. - Korkeakoulututkimuksen kulttuurinäkökulma. 94 p. Yhteenveto 5 p. 1995.
- 114 KAIPIO, KALEVI, Yhteisöllisyys kasvatuksessa. yhteisökasvatuksen teoreettinen analyysi ja käytäntöön soveltaminen. - The community as an educator. Theoretical analysis and practice of community education. 250 p. Summary 3 p. 1995.
- 115 HÄNNIKÄINEN, MARITTA, Nukesta vauvaksi ja lapsesta lääkäriksi. Roolileikkiin siirtymisen tarkastelua piagetilaisesta ja kulttuurihistoriallisen toiminnan teorian näkökulmasta. 73 p. Summary 6 p. 1995.
- 116 IKONEN, OIVA. Adaptiivinen opetus. Oppimistutkimus harjaantumiskoulun opetussuunnitelma- ja seurantajärjestelmän kehittämisen tukena. - The adaptive teaching. 90 p. Summary 5 p. 1995.
- 117 SUUTAMA, TIMO. Coping with life events in old age. - Elämän muutos- ja ongelmatilanteiden käsittely iäkkäillä ihmisillä. 110 p. Yhteenveto 3 p. 1995.
- 118 DERSEH, TIBEBU BOGALE, Meanings Attached to Disability, Attitudes towards Disabled People, and Attitudes towards Integration. 150 p. 1995.
- 119 SAHLBERG, PASI, Kuka auttaisi opettajaa. Post-moderni näkökulma opetuksen muu-tokseen yhden kehittämisprojektin valossa. - Who would help a teacher. A post-modern perspective on change in teaching in light of a school improvement project. 255 p. Summary 4 p. 1996.
- 120 UHINKI, AILO, Distress of unemployed job-seekers described by the Zulliger Test using the Comprehensive System. - Työttömien työntekijöiden ahdinko kuvattuna Comprehensive Systemin mukaisesti käytetyillä Zulligerin testillä. 61 p. Yhteenveto 3p. 1996.
- 121 ANTIKAINEN, RISTO, Clinical course, outcome and follow-up of inpatients with borderline level disorders. - Rajatilapotilaiden osastohoidon tuloksellisuus kolmen vuoden seurantatutkimuksessa Kys:n psykiatrian klinikassa. 102 p. Yhteenveto 4 p. 1996.
- 122 RUUSUVIRTA, TIMO, Brain responses to pitch changes in an acoustic environment in cats and rabbits. - Aivovasteet kuuloärsykemuutoksiin kissoilla ja kaneilla. 45 p. Yhteenveto 2 p. 1996.
- 123 VISTI, ANNALISA, Työyhteisön ja työn tuottavuuden kehitys organisaation transformaatiossa. - Development of the work community and changes in the productivity of work during an organizational transformation process. 201 p. Summary 12 p. 1996.
- 124 SALLINEN, MIKAEL, Event-related brain potentials to changes in the acoustic environment during sleep and sleepiness. - Aivojen heränevasteet muutoksiin kuuloärsykesarjassa unen ja uneliaisuuden aikana. 104 p. Yhteenveto 3 p. 1997.
- 125 LAMMINMÄKI, TUUJA, Efficacy of a multi-faceted treatment for children with learning difficulties. - Oppimisvaikeuksien neurokognitiivisen ryhmäkuntoutuksen tuloksellisuus ja siihen vaikuttavia tekijöitä. 56 p. Yhteenveto 2 p. 1997.
- 126 LUTINEN, JAANA, Fragmentoituva kulttuuripolitiikka. Paikallisen kulttuuripolitiikan tulkintakehykset Ylä-Savossa. - Fragmenting-cultural policy. The interpretative frames of local cultural politics in Ylä-Savo. 178 p. Summary 9 p. 1997.
- 127 MARTTUNEN, MIIKA, Studying argumentation in higher education by electronic mail. - Argumentointia yliopisto-opinnoissa sähköpostilla. 60 p. (164 p.) Yhteenveto 3 p. 1997.
- 128 JAAKKOLA, HANNA, Kielitieto kielitaitoon pyrittäessä. Vieraiden kielten opettajien käsityksiä kieliopin oppimisesta ja opettamisesta. - Language knowledge and language ability. Teachers' conceptions of the role of grammar in foreign language learning and teaching. 227 p. Summary 7 p. 1997.
- 129 SUBRA, LEENA, A portrait of the political agent in Jean-Paul Sartre. Views on playing, acting, temporality and subjectivity. - Poliittisen toimijan muotokuva Jean-Paul Sartrella. Näkymiä pelaamiseen, toimintaan, ajallisuuteen ja subjektiivisuuteen. 248 p. Yhteenveto 2 p. 1997.
- 130 HAARAKANGAS, KAUKO, Hoitokokouksen äänet. Dialoginen analyysi perhekeskeisen psykiatrisen hoitoprosessin hoitokokous-keskusteluita työryhmän toiminnan näkökulmasta. - The voices in treatment meeting. A dialogical analysis of the treatment meeting conversations in family-centred psychiatric treatment process in regard to the team activity. 136 p. Summary 8 p. 1997.
- 131 MATINHEIKKI-KOKKO, KAIJA, Challenges of working in a cross-cultural environment. Principles and practice of refugee settlement in Finland. - Kulttuurienvälisen työn haasteet. Periaatteet ja käytäntö maahanmuuttajien hyvinvoinnin turvaamiseksi Suomessa. 130 p. Yhteenveto 3 p. 1997.
- 132 KIVINIEMI, KARI, Opettajuuden oppimisesta harjoittelun harhautuksiin. Aikuisopiskelijoiden kokemuksia opetusharjoittelusta ja sen ohjauksesta luokanopettajakoulutuksessa. - From the learning of teacherhood to the fabrications of practice. Adult students' experiences of teaching practice and its supervision in class teacher education. 267 p. Summary 8 p. 1997.
- 133 KANTOLA, JOUKO, Cygnaeuksen jäljillä käsityönopetuksesta teknologiseen kasvatukseen. - In the footsteps of Cygnaeus. From handicraft teaching to technological education. 211 p. Summary 7 p. 1997.
- 134 KAARTINEN, JUKKA, Nocturnal body movements

- and sleep quality. - Yölliset kehon liikkeet ja unen laatu. 85 p. Yhteenveto 3 p. 1997.
- 135 MUSTONEN, ANU, Media violence and its audience. - Mediaväkivalta ja sen yleisö. 44 p. (131 p.). Yhteenveto 2 p. 1997.
- 136 PERTTULA, JUHA, The experienced life-fabrics of young men. - Nuorten miesten koettu elämänkudelman. 218 p. Yhteenveto 3 p. 1998.
- 137 TIKKANEN, TARJA, Learning and education of older workers. Lifelong learning at the margin. - Ikääntyvän työväestön oppiminen ja koulutus. Elinikäisen oppimisen marginaalissa. 83 p. (154 p.). Yhteenveto 6 p. 1998.
- 138 LEINONEN, MARKKU, Johannes Gezelius vanhempi luonnonmukaisen pedagogiikan soveltajana. Comeniuslainen tulkinta. - Johannes Gezelius the elder as implementer of natural pedagogy. A Comenian interpretation. 237 p. Summary 7 p. 1998.
- 139 KALLIO, EEVA, Training of students' scientific reasoning skills. - Korkeakouluopiskelijoiden tieteellisen ajattelun kehittäminen. 90 p. Yhteenveto 1 p. 1998.
- 140 NIEMI-VÄKEVÄINEN, LEENA, Koulutusjaksot ja elämänpolitiikka. Koulutautuminen yksilöllistymisen ja yhteisöllisyyden risteysasemana. - Sequences of vocational education as life politics. Perspectives of individualization and communality. 210 p. Summary 6 p. 1998.
- 141 PARIKKA, MATTI, Teknologiaкомпетенssi. Teknologiakasvatuksen uudistamishaasteita peruskoulussa ja lukiossa. - Technological competence. Challenges of reforming technology education in the Finnish comprehensive and upper secondary school. 207 p. Summary 13 p. 1998.
- 142 TA OPETTAJAN APUNA - EDUCATIONAL TA FOR TEACHER. Professori Pirkko Liikaselle omistettu juhla-kirja. 207 p. Tiivistelmä - Abstract 14 p. 1998.
- 143 YLÖNEN, HILKKA, Taikahattu ja hopeakengät - sadun maailmaa. Lapsi päiväkodissa sadun kuulijana, näkijänä ja kokijana. - The world of the colden cap and silver shoes. How kindergarten children listen to, view, and experience fairy tales. 189 p. Summary 8 p. 1998.
- 144 MOILANEN, PENNITTI, Opettajan toiminnan perusteiden tulkinta ja tulkinnan totuudellisuuden arviointi. - Interpreting reasons for teachers' action and the verifying the interpretations. 226 p. Summary 3p. 1998.
- 145 VAURIO, LEENA, Lexical inferencing in reading in english on the secondary level. - Sanapäätö englanninkielistä tekstiä luettaessa lukioasteella. 147 p. Yhteenveto 3 p. 1998.
- 146 ETELÄPELTO, ANNELI, The development of expertise in information systems design. - Asiantuntijuuden kehittyminen tietojärjestelmien suunnittelussa. 132 p. (221p.). Yhteenveto 12 p. 1998.
- 147 PIIRHONEN, ANTTI, Redundancy as a criterion for multimodal user-interfaces. - Käsitteistö luo näkökulman käyttöliittymäanalyyysiin. 141 p. Yhteenveto 3 p. 1998.
- 148 RÖNKÄ, ANNA, The accumulation of problems of social functioning: outer, inner, and behavioral strands. - Sosiaalinen selviytyminen lapsuudesta aikuisuuteen: ongelmien kasautumisen kolme väylää. 44 p. (129 p.) Yhteenveto 3 p. 1999.
- 149 NAUKKARINEN, AIMO, Tasapainoilua kurinalaisuuden ja tarkoituksenmukaisuuden välillä. Oppilaiden ei-toivottuun käyttäytymiseen liittyvän ongelmanratkaisun kehittäminen yhden peruskoulun yläasteen tarkastelun pohjalta. - Balancing rigor and relevance. Developing problem-solving associated with students' challenging behavior in the light of a study of an upper comprehensive school. 296 p. Summary 5 p. 1999.
- 150 HOLMA, JUHA, The search for a narrative. Investigating acute psychosis and the need-adapted treatment model from the narrative viewpoint. - Narratiivinen lähestymistapa akuuttiin psykoosiin ja tarpeenmukaisen hoidon malliin. 52 p. (105 p.) Yhteenveto 3 p. 1999.
- 151 LEPPÄNEN, PAAVO H.T., Brain responses to changes in tone and speech stimuli in infants with and without a risk for familial dyslexia. - Aivovasteet ääni- ja puheärsykkeiden muutoksiin vauvoilla, joilla on riski suvussa esiintyvään dysleksiaan ja vauvoilla ilman tätä riskiä. 100 p. (197 p.) Yhteenveto 4 p. 1999.
- 152 SUOMALA, JYRKI, Students' problem solving in the LEGO/Logo learning environment. - Oppilaiden ongelmanratkaisu LEGO/Logo oppimisympäristössä. 146 p. Yhteenveto 3 p. 1999.
- 153 HUTTUNEN, RAUNO, Opettämisen filosofia ja kritiikki. - Philosophy, teaching, and critique. Towards a critical theory of the philosophy of education. 201 p. Summary 3p. 1999.
- 154 KAREKIVI, LEENA, Ehkä en kokeilisikaan, jos Tutkimus ylivieskalaisten nuorten tupakoinnista ja päihteidenkäytöstä ja niihin liittyvästä terveyskasvatuksesta vuosina 1989-1998. - Maybe I wouldn't even experiment if A study on youth smoking and use of intoxicants in Ylivieska and related health education in 1989-1998. 256 p. Summary 4 p. 1999.
- 155 LAAKSO, MARJA-LEENA, Prelinguistic skills and early interactional context as predictors of children's language development. - Esi-kielellinen kommunikaatio ja sen vuorovaikutuksellinen konteksti lapsen kielen kehityksen ennustajana. 127 p. Yhteenveto 2 p. 1999.
- 156 MAUNO, SAIIJA, Job insecurity as a psycho-social job stressor in the context of the work-family interface. - Työn epävarmuus työn psykososiaalisena stressitekijänä työn ja perheen vuorovaikutuksen kontekstissa. 59 p. (147 p.) Yhteenveto 3 p. 1999.
- 157 MÄENSIVU KIRSTI, Opettaja määrittelijänä, oppilas määriteltävänä. Sanallisen oppilaan arvioinnin sisällön analyysi. - The teacher as a determiner - the pupil to be determined -

- content analysis of the written school reports. 215 p. Summary 5 p. 1999.
- 158 FELDT, TARU, Sense of coherence. Structure, stability and health promoting role in working life. - Koherenssin rakenne, pysyvyys ja terveyttä edistävä merkitys työelämässä. 60 p. (150 p.) Yhteenveto 5 p. 2000.
- 159 MÄNTY, TARJA, Ammatillisista erityisoppilaitoksista elämään. - Life after vocational special education. 235 p. Summary 3 p. 2000.
- 160 SARJA, ANNELI, Dialogioppiminen pienryhmässä. Opettajaksi opiskelevien harjoitteluprosessi terveydenhuollon opettajankoulutuksessa. - Dialogic learning in a small group. The process of student teachers' teaching practice during health care education. 165 p. Summary 7 p. 2000.
- 161 JÄRVINEN, ANITTA, Taitajat iänikuiset. - Kotkan ammattilukiosta valmiuksia elämään, työelämään ja jatko-opintoihin. - Age-old craftmasters -Kotka vocational senior secondary school - giving skills for life, work and further studies. 224 p. Summary 2 p. 2000.
- 162 KONTIO, MARJA-LIISA, Laitoksessa asuvan kehitysvammaisen vanhuksen haastava käyttäytyminen ja hoitajan käyttämiä vaikutuskeinoja. - Challenging behaviour of institutionalized mentally retarded elderly people and measures taken by nurses to control it. 175 p. Summary 3 p. 2000.
- 163 KILPELÄINEN, ARJA, Naiset paikkaansa etsimässä. Aikuiskoulutus naisen elämänsä rakentajana. - Adult education as determinant of woman's life-course. 155 p. Summary 6 p. 2000.
- 164 RIITESUO, ANNIKKI, A preterm child grows. Focus on speech and language during the first two years. - Keskonen kasvaa: puheen ja kielen kehitys kahtena ensimmäisenä elinvuotena. 119 p. Tiivistelmä 2 p. 2000.
- 165 TAURIAINEN, LEENA, Kohti yhteistä laatua. - Henkilökunnan, vanhempien ja lasten laatu-käsitykset päiväkodin integroidussa erityisryhmässä. - Towards common quality: staff's, parents' and children's conceptions of quality in an integration group at a daycare center. 256 p. Summary 6 p. 2000.
- 166 RAUDASKOSKI, LEENA, Ammattikorkeakoulun toimintaperustaa etsimässä. Toimilupahakemusten sisällönanalyttinen tarkastelu. - In search for the founding principles of the Finnish polytechnic institutes. A content analysis of the licence applications. 193 p. Summary 4 p. 2000.
- 167 TAKKINEN, SANNA, Meaning in life and its relation to functioning in old age. - Elämän tarkoituksellisuus ja sen yhteydet toimintakykyyn vanhuudessa. 51 p. (130 p.) Yhteenveto 2 p. 2000.
- 168 LAUNONEN, LEEVI, Eettinen kasvatusajattelu suomalaisen koulun pedagogisissa teksteissä 1860-luvulta 1990-luvulle. - Ethical thinking in Finnish school's pedagogical texts from the 1860s to the 1990s. 366 p. Summary 3 p. 2000.
- 169 KUORELAHTI, MATTI, Sopeutumattomien luokkamutoisen erityisopetuksen tuloksellisuus. - The educational outcomes of special classes for emotionally/ behaviorally disordered children and youth. 176 p. Summary 2p. 2000.
- 170 KURUNMÄKI, JUSSI, Representation, nation and time. The political rhetoric of the 1866 parliamentary reform in Sweden. - Edustus, kansakunta ja aika. Poliittinen retoriikka Ruotsin vuoden 1866 valtiopäiväreformissa. 253 p. Tiivistelmä 4 p. 2000.
- 171 RASINEN, AKI, Developing technology education. In search of curriculum elements for Finnish general education schools. 158 p. Yhteenveto 2 p. 2000.
- 172 SUNDHOLM, LARS, Itseohjautuvuus organisaatiomuutoksessa. - Self-determination in organisational change. 180 p. Summary 15 p. 2000.
- 173 AHONNISKKA-ASSA, JAANA, Analyzing change in repeated neuropsychological assessment. 68 p. (124 p.) Yhteenveto 2 p. 2000.
- 174 HOFFRÉN, JARI, Demokraattinen eetos - rajoista mahdollisuuksiin. - The democratic ethos. From limits to possibilities? 217 p. Summary 2 p. 2000.
- 175 HEIKKINEN, HANNU L. T., Toimintatutkimus, tarinat ja opettajaksi tuleminen taito. Narratiivisen identiteettityön kehittäminen opettajankoulutuksessa toimintatutkimuksen avulla. - Action research, narratives and the art of becoming a teacher. Developing narrative identity work in teacher education through action research. 237 p. Summary 4 p. 2001.
- 176 VUORENMAA, MARITTA, Ikkunoita arvioinnin tuolle puolen. Uusia avauksia suomalaisen koulutusta koskevaan evaluaatiokeskusteluun. - Views across assessment: New openings into the evaluation discussion on Finnish education. 266 p. Summary 4 p. 2001.
- 177 LITMANEN, TAPIO, The struggle over risk. The spatial, temporal, and cultural dimensions of protest against nuclear technology. - Kampailu riskistä. Ydinteknologian vastaisen protestin tilalliset, ajalliset ja kulttuuriset ulottuvuudet. 72 p. (153 p.) Yhteenveto 9 p. 2001.
- 178 AUNOLA, KAISA, Children's and adolescents' achievement strategies, school adjustment, and family environment. - Lasten ja nuorten suoritusstrategiat koulu- ja perheympäristöissä. 51 p. (153 p.) Yhteenveto 2 p. 2001.
- 179 OKSANEN, ELINA, Arvioinnin kehittäminen erityisopetuksessa. Diagnosoinnista oppimisen ohjaukseen laadullisena tapaustutkimuksena. - Developing assessment practices in special education. From a static approach to

- dynamic approach applying qualitative case. 182 p. Summary 5 p. 2001.
- 180 VIITTALA, KAISU, "Kyllä se tommosellaki lapsella on kovempi urakka". Sikiöaikana alkoholille altistuneiden huostaanotettujen lasten elämäntilanne, riskiprosessit ja suojaavat prosessit. - "It's harder for that kind of child to get along". The life situation of the children exposed to alcohol in utero and taken care of by society, their risk and protective processes. 316 p. Summary 4 p. 2001.
- 181 HANSSON, LEENI, Networks matter. The role of informal social networks in the period of socioeconomic reforms of the 1990s in Estonia. - Verkostoilla on merkitystä: infor-maalisten sosiaalisten verkostojen asema Virossa 1990-luvun sosio-ekonomisten muutosten aikana. 194 p. Yhteenveto 3 p. 2001.
- 182 BÖÖK, MARJA LEENA, Vanhemmuus ja vanhemmuuden diskurssit työttömyystilanteessa. - Parenthood and parenting discourses in a situation of unemployment. 157 p. Summary 5 p. 2001.
- 183 KOKKO, KATJA, Antecedents and consequences of long-term unemployment. - Pitkäaikaistyöttömyyden ennakoijia ja seurauksia. 53 p. (115 p.) Tiivistelmä 3 p. 2001.
- 184 KOKKONEN, MARJA, Emotion regulation and physical health in adulthood: A longitudinal, personality-oriented approach. - Aikuisiän tunteiden säätely ja fyysinen terveys: pitkittäistutkimuksellinen ja persoonallisuuskeskeinen lähestymistapa. 52 p. (137 p.) Tiivistelmä 3 p. 2001.
- 185 MÄNNIKKÖ, KAISA, Adult attachment styles: A Person-oriented approach. - Aikuisten kiintymystyyli. 142 p. Yhteenveto 5 p. 2001.
- 186 KATVALA, SATU, Missä äiti on? Äitejä ja äitiyden uskomuksia sukupolvien saatossa. - Where's mother? Mothers and maternal beliefs over generations. 126 p. Summary 3 p. 2001.
- 187 KIISKINEN, ANNA-LIISA, Ympäristöhallinto vastuullisen elämäntavan edistäjänä. - Environmental administration as promoter of responsible living. 229 p. Summary 8 p. 2001.
- 188 SIMOLA, AHTI, Työterveyshuolto-organisaation toiminta, sen henkilöstön henkinen hyvinvointi ja toiminnan tuloksellisuus. - Functioning of an occupational health service organization and its relationship to the mental well-being of its personnel, client satisfaction, and economic profitability. 192 p. Summary 12 p. 2001.
- 189 VESTERINEN, PIRKKO, Projektiopiskelu- ja oppiminen ammattikorkeakoulussa. - Project-based studying and learning in the polytechnic. 257 p. Summary 5 p. 2001.
- 190 KEMPPAINEN, JAANA, Kotikasvatus kolmessa sukupolvessa. - Childrearing in three generations. 183 p. Summary 3 p. 2001.
- 191 HÖHENTHAL-ANTIN LEONIE, Luvan ottaminen - lkäihmiset teatterin tekijöinä. - Taking permission- Elderly people as theatre makers. 183 p. Summary 5 p. 2001.
- 192 KAKKORI, LEENA, Heideggerin aukeama. Tutkimuksia totuudesta ja taiteesta Martin Heideggerin avaamassa horisontissa. - Heidegger's clearing. Studies on truth and art in the horizon opened by Martin Heidegger. 156 p. Summary 2 p. 2001.
- 193 NÄRHI, VESA, The use of clinical neuropsychological data in learning disability research. - Asiakastyön yhteydessä kerätyn neuropsykologisen aineiston käyttö oppimisvaikeustutkimuksessa. 103 p. Yhteenveto 2 p. 2002.
- 194 SUOMI, ASTA, Ammattia etsimässä. Aikuisopiskelijat kertovat sosiaaliohjaajakoulutuksesta ja narratiivisen pätevyyden kehittymisestä. - Searching for professional identity. Adult students' narratives on the education of a social welfare supervisor and the development of narrative competence. 183 p. Summary 2 p. 2002.
- 195 PERKKILÄ, PÄIVI, Opettajien matematiikkauskomukset ja matematiikan oppikirjan merkitys alkuopetuksessa. 212 p. - Teacher's mathematics beliefs and meaning of mathematics textbooks in the first and the second grade in primary school. Summary 2 p. 2002.
- 196 VESTERINEN, MARJA-LIISA, Ammatillinen harjoittelu osana asiantuntijuuden kehittymistä ammattikorkeakoulussa. - Promoting professional expertise by developing practical learning at the polytechnic. 261 p. Summary 5 p. 2002.
- 197 POHJANEN, JORMA, Mitä kello on? Kello modernissa yhteiskunnassa ja sen sosiologisessa teoriassa. - What's the time. Clock on modern society and in it's sociological theory. 226 p. Summary 3 p. 2002.
- 198 RANTALA, ANJA, Perhekeskeisyys - puhetta vai todellisuutta? Työntekijöiden käsitykset yhteistyöstä erityistä tukea tarvitsevan lapsen perheen kanssa. - Family-centeredness rhetoric or reality? Summary 3 p. 2002.
- 199 VALANNE, EIJA, "Meidän lapsi on arvokas" Henkilökohtainen opetuksen järjestämistä koskeva suunnitelma (HOJKS) kunnallisessa erityiskoulussa. - "Our child is precious" - The individual educational plan in the context of the special school. 219 p. Yhteenveto 2 p. 2002.
- 200 HOLOPAINEN, LEENA, Development in reading and reading related skills; a follow-up study from pre-school to the fourth grade. 57 p. (138 p.) Yhteenveto 3 p. 2002.
- 201 HEIKKINEN, HANNU, Draaman maailmat oppimisalueina. Draamakasvatuksen vakava leikillisuus. - Drama worlds as learning areas - the serious playfulness as drama education. 164 p. Summary 5 p. 2002.
- 202 HYTÖNEN, TUJJA, Exploring the practice of human resource development as a field of professional expertise. - Henkilöstön

- kehittämistyön asiantuntijuuden rakentumisen. 137 p. (300 p.) Yhteenveto 10 p. 2002.
- 203 RIPATTI, MIKKO, Arvid Järnefeldt kasvatusajattelijana. 246 p. Summary 4 p. 2002.
- 204 VIRMASALO, ILKKA, Perhe, työttömyys ja lama. - Families, unemployment and the economic depression. 121 p. Summary 2 p. 2002.
- 205 WIKGREN, JAN, Diffuse and discrete associations in aversive classical conditioning. - Täsmälliset ja laaja-alaiset ehdollistumat klassisessa aversiivisessä ehdollistumisessa. 40 p. (81 p.) Yhteenveto 2 p. 2002.
- 206 JOKIVUORI, PERTTI, Sitoutuminen työorganisaatioon ja ammattijärjestöön. - Kilpailevia vai täydentäviä? - Commitment to organisation and trade union. Competing or complementary? 132 p. Summary 8 p. 2002.
- 207 GONZÁLEZ VEGA, NARCISO, Factors affecting simulator-training effectiveness. 162 p. Yhteenveto 1 p. 2002.
- 208 SALO, KARI, Teacher Stress as a Longitudinal Process - Opettajien stressiprosessi. 67 p. Yhteenveto 2 p. 2002.
- 209 VAUHKONEN, JOUNI, A rhetoric of reduction. Bertrand de Jouvenel's pure theory of politics as persuasion. 156 p. Tiivistelmä 2 p. 2002.
- 210 KONTONIEMI, MARITA, "Milloin sinä otat itseäsi niskasta kiinni?" Opettajien kokemuksia alisuoriutujista. - "When will you pull your socks up?" Teachers' experiences of underachievers. 218 p. Summary 3 p. 2003.
- 211 SAUKKONEN, SAKARI, Koulu ja yksilöllisyys; Jännitteitä, haasteita ja mahdollisuuksia. - School and individuality: Tensions, challenges and possibilities. 125 p. Summary 3 p. 2003.
- 212 VILJAMAA, MARJA-LEENA, Neuvola tänään ja huomenna. Vanhemmuuden tukeminen, perhekeskeisyys ja vertaistuki. - Child and maternity welfare clinics today and tomorrow. Supporting parenthood, family-centered services and peer groups. 141 p. Summary 4 p. 2003.
- 213 REMES, LIISA, Yrittäjyyskasvatuksen kolme diskurssia. - Three discourses in entrepreneurial learning. 204 p. Summary 2 p. 2003.
- 214 KARJALA, KALLE, Neulanreiästä panoraamaksi. Ruotsin kulttuurikuvan ainekset eräissä keskikoulun ja B-ruotsin vuosina 1961-2002 painetuissa oppikirjoissa. - From pinhole to panorama - The culture of Sweden presented in some middle and comprehensive school textbooks printed between 1961 and 2002. 308 p. Summary 2 p. 2003.
- 215 LALLUKKA, KIRSI, Lapsuusikä ja ikä lapsuudessa. Tutkimus 6-12-vuotiaiden sosiokulttuurisesta ikätiedosta. - Childhood age and age in childhood. A study on the sociocultural knowledge of age. 234 p. Summary 2 p. 2003.
- 216 PUUKARI, SAULL, Video Programmes as Learning Tools. Teaching the Gas Laws and Behaviour of Gases in Finnish and Canadian Senior High Schools. 361 p. Yhteenveto 6 p. 2003.
- 217 LOISA, RAIJA-LEENA, The polysemous contemporary concept. The rhetoric of the cultural industry. - Monimerkityksinen nykykäsite. Kulttuuriteollisuuden retoriikka. 244 p. Yhteenveto 2 p. 2003.
- 218 HOLOPAINEN, ESKO, Kuullun ja luetun tekstin ymmärtämisstrategiat ja -vaikeudet peruskoulun kolmannella ja yhdeksännellä luokalla. - Strategies for listening and reading comprehension and problematic listening and reading comprehension of the text during the third and ninth grades of primary school. 135 p. Summary 3 p. 2003.
- 219 PENTTINEN, SEPPÖ, Lähtökohdat liikuntaa opettavaksi luokanopettajaksi. Nuoruuden kasvuympäristöt ja opettajankoulutus opettajuuden kehitystekijöinä. - Starting points for a primary school physical education teacher. The growth environment of adolescence and teacher education as developmental factors of teachership. 201 p. Summary 10 p. 2003.
- 220 IKÄHEIMO, HEIKKI, Tunnustus, subjektiviteetti ja inhimillinen elämänmuoto: Tutkimuksia Hegelistä ja persoonien välisistä tunnustus-suhteista. - Recognition, subjectivity and the human life form: studies on Hegel and interpersonal recognition. 191 p. Summary 3 p. 2003.
- 221 ASUNTA, TUULA, Knowledge of environmental issues. Where pupils acquire information and how it affects their attitudes, opinions, and laboratory behaviour - Ympäristöasioita koskeva tieto. Mistä oppilaat saavat informaatiota ja miten se vaikuttaa heidän asenteisiinsa, mielipiteisiinsä ja laboratoriokäyttäytymiseensä. 159 p. Yhteenveto 4 p. 2003.
- 222 KUJALA, ERKKI, Sodan pojat. Sodanaikaisten pikkupoikien lapsuuskokemuksia isyyden näkökulmasta - The sons of war. 229 p. Summary 2 p. 2003.
- 223 JUSSI KURUNMÄKI & KARI PALOINEN (Hg./eds.) Zeit, Geschichte und Politik. Time, history and politics. *Zum achtzigsten Geburtstag von Reinhart Koselleck*. 310 p. 2003.
- 224 LAITINEN, ARTO, Strong evaluation without sources. On Charles Taylor's philosophical anthropology and cultural moral realism. - Vahvoja arvostuksia ilman lähteitä. Charles Taylorin filosofisesta antropologiasta ja kulturalistisesta moraalirealismista. 358 p. Yhteenveto 4 p. 2003.
- 225 GUTTORM, TOMI K. Newborn brain responses measuring feature and change detection and predicting later language development in children with and without familial risk for dyslexia. - Vastasyntyneiden aivovasteet puheäänteiden ja niiden muutosten havaitsemisessa sekä myöhemmän kielen kehityksen ennustamisessa dysleksia-riskilapsilla. 81 p. (161 p.) Yhteenveto 3 p. 2003.

- 226 NAKARI, MAIJA-LIISA, Työilmapiiri, työntekijöiden hyvinvointi ja muutoksen mahdollisuus - Work climate, employees' well-being and the possibility of change. 255 p. Summary 3 p. 2003.
- 227 METSÄPELTO, RIITTA-LEENA, Individual differences in parenting: The five-factor model of personality as an explanatory framework - Lastenkasvatus ja sen yhteys vanhemman persoonallisuuden piirteisiin. 53 p. (119 p.) Tiivistelmä 3 p. 2003.
- 228 PULKKINEN, OILI, The labyrinth of politics - A conceptual approach to the modes of the political in the scottish enlightenment. 144 p. Yhteenveto 2 p. 2003.
- 229 JUUJARVI, PETRI, A three-level analysis of reactive aggression among children. - Lasten aggressiivisiin puolustusreaktioihin vaikuttavien tekijöiden kolmitasoinen analyysi. 39 p. (115 p.) Yhteenveto 2 p. 2003.
- 230 POIKONEN, PIIRJO-LIISA, "Opetussuunnitelma on sitä elämää". Päiväkoti-koulu yhteisö opetussuunnitelman kehittäjänä. - "The curriculum is part of our life". The day-care - primary school community as a curriculum developer. 154 p. Summary 3 p. 2003.
- 231 SOININEN, SUVI, From a 'Necessary Evil' to an art of contingency: Michael Oakeshott's conception of political activity in British postwar political thought. 174 p. Summary 2p. 2003.
- 232 ALARAUDANJOKI, ESA, Nepalese child labourers' life-contexts, cognitive skills and well-being. - Työssäkäyvien nepalilaislasten elämänkonteksti, kognitiiviset taidot ja hyvinvointi. 62 p. (131 p.) Yhteenveto 4 p. 2003.
- 233 LERKKANEN, MARJA-KRISTINA, Learning to read. Reciprocal processes and individual pathways. - Lukemaan oppiminen: vastavuoroiset prosessit ja yksilölliset oppimispolut. 70 p. (155 p.) Yhteenveto 5 p. 2003.
- 234 FRIMAN, MERVI, Ammatillisen asiantuntijan etiikka ammattikorkeakoulutuksessa. - The ethics of a professional expert in the context of polytechnics. 199 p. 2004.
- 235 MERONEN, AULI, Viittomakielen omaksumisen yksilölliset tekijät. - Individual differences in sign language abilities. 110 p. Summary 5 p. 2004.
- 236 TIILIKKALA, LIISA, Mestarista tuutoriksi. Suomalaisen ammatillisen opettajuuden muutos ja jatkuvuus. - From master to tutor. Change and continuity in Finnish vocational teacherhood. 281 p. Summary 3 p. 2004.
- 237 ARO, MIKKO, Learning to read: The effect of orthography. - Kirjoitusjärjestelmän vaikutus lukemaan oppimiseen. 44 p. (122 p.) Tiivistelmä 2 p. 2004.
- 238 LAAKSO, ERKKI, Draamakokemusten äärellä. Prosessidraaman oppimispotentiaali opettajaksi opiskelevien kokemusten valossa. - Encountering drama experiences. The learning potential of process drama in the light of student teachers' experiences. 230 p. Summary 7 p. 2004.
- 239 PERÄLÄ-LITTUNEN, SATU, Cultural images of a good mother and a good father in three generations. - Kulttuuriset mielikuvat hyvästä äidistä ja hyvästä isästä kolmessa sukupolvessa. 234 p. Yhteenveto 7 p. 2004.
- 240 RINNE-KOISTINEN, EVA-MARITA, Perceptions of health: Water and sanitation problems in rural and urban communities in Nigeria. 129 p. (198 p.) Yhteenveto 3 p. 2004.
- 241 PALMROTH, AINO, Käännösten kautta kollektiiviin. Tuuliosuuskunnat toimijaverkkoina. - From translation to collective. Wind turbine cooperatives as actor networks. 177 p. Summary 7 p. 2004.
- 242 VIERIKKO, ELINA, Genetic and environmental effects on aggression. - Geneettiset ja ympäristötekijät aggressiivisuudessa. 46 p. (108 p.) Tiivistelmä 3 p. 2004.
- 243 NÄRHI, KATI, The eco-social approach in social work and the challenges to the expertise of social work. - Ekososiaalinen viitekehys ja haasteet sosiaalityön asiantuntijuudelle. 106 p. (236 p.) Yhteenveto 7 p. 2004.
- 244 URSIN, JANI, Characteristics of Finnish medical and engineering research group work. - Tutkimusryhmätyöskentelyn piirteet lääke- ja teknisissä tieteissä. 202 p. Yhteenveto 9 p. 2004.
- 245 TREUTHARDT, LEENA, Tulostehokkuuden yhteiskunnallisuus Jyväskylän yliopistossa. Tarkastelunäkökulmina muotit ja seurustelu. - The management by results a fashion and social interaction at the University of Jyväskylä. 228 p. Summary 3 p. 2004.
- 246 MATTHIES, JÜRGEN, Umweltpädagogik in der Postmoderne. Eine philosophische Studie über die Krise des Subjekts im umweltpädagogischen Diskurs. - Ympäristökasvatus postmodernissa. Filosofinen tutkimus subjektin kriisistä ympäristökasvatuksen diskurssissa. 400 p. Yhteenveto 7 p. 2004.
- 247 LAITILA, AARNO, Dimensions of expertise in family therapeutic process. - Asiantuntijuuden ulottuvuuksia perheterapeuttisessa prosessissa. 54 p. (106 p.) Yhteenveto 3 p. 2004.
- 248 LAAMANEN (ASTIKAINEN), PIIA, Pre-attentive detection of changes in serially presented stimuli in rabbits and humans. - Muutoksen esitietoinen havaitseminen sarjallisesti esitetyissä ärsykkeissä kaneilla ja ihmisillä. 35 p. (54 p.) Yhteenveto 3 p. 2004.
- 249 JUUSENAHO, RIITTA, Peruskoulun rehtoreiden johtamisen eroja. Sukupuolinen näkökulma. - Differences in comprehensive school leadership and management. A gender-based approach. 176p. Summary 3 p. 2004.

- 250 VAARAKALLIO, TUULA, "Rotten to the Core". Variations of French nationalist anti-system rhetoric. - "Systeemi on mätä". Ranskalaisien nationalistien järjestelmän vastainen retoriikka. 194 p. Yhteenveto 3 p. 2004.
- 251 KUUSINEN, PATRIK, Pitkäaikainen kipu ja depressio. Yhteyttä säätelevät tekijät. - Chronic pain and depression: psychosocial determinants regulating the relationship. 139 p. Summary 8 p. 2004.
- 252 HÄNNIKÄINEN-UUTELA, ANNA-LIISA, Uudelleen juurtuneet. Yhteisökasvatus vaikeasti päihderiippuvaisten narkomaanien kuntoutuksessa. - Rooted again. Community education in the rehabilitation of substance addicts. 286 p. Summary 3 p. 2004.
- 253 PALONIEMI, SUSANNA, Ikä, kokemus ja osaaminen työelämässä. Työntekijöiden käsityksiä iän ja kokemuksen merkityksestä ammatillisessa osaamisessa ja sen kehittämisessä. - Age, experience and competence in working life. Employees' conceptions of the the meaning and experience in professional competence and its development. 184 p. Summary 5 p. 2004.
- 254 RUIZ CEREZO, MONTSE, Anger and Optimal Performance in Karate. An Application of the IZOF Model. 55 p. (130 p.) Tiivistelmä 2 p. 2004.
- 255 LADONLAHTI, TARJA, Haasteita palvelujärjestelmälle. Kehitysvammaiseksi luokiteltu henkilö psykiatrisessa sairaalassa. - Challenges for the human service system. Living in a psychiatric hospital under the label of mental retardation. 176 p. Summary 3 p. 2004.
- 256 KOVANEN PÄIVI, Oppiminen ja asiantuntijuus varhaiskasvatuksessa. Varhaisen oppimaan ohjaamisen suunnitelma erityistä tukea tarvitsevien lasten ohjauksessa. - Learning and expertise in early childhood education. A pilot work in using VARSU with children with special needs. 175 p. Summary 2 p. 2004.
- 257 VILMI, VEIKKO, Turvallinen koulu. Suomalaisien näkemyksiä koulutuspalvelujen kansallisesta ja kunnallisesta priorisoinnista. - Secure education. Finnish views on the national and municipal priorities of Finland's education services. 134 p. Summary 5 p. 2005.
- 258 ANTTILA, TIMO, Reduced working hours. Reshaping the duration, timing and tempo of work. 168 p. Tiivistelmä 2 p. 2005.
- 259 UGASTE, AINO, The child's play world at home and the mother's role in the play. 207 p. Tiivistelmä 5 p. 2005.
- 260 KURRI, KATJA, The invisible moral order: Agency, accountability and responsibility in therapy talk. 38 p. (103 p.). Tiivistelmä 1 p. 2005.
- 261 COLLIN, KAIJA, Experience and shared practice - Design engineers' learning at work. - Suunnitteluinsinöörien työssä oppiminen - kokemuksellisuutta ja jaettuja käytäntöjä. 124 p. (211 p.). Yhteenveto 6 p. 2005.
- 262 KURKI, EIJA, Näkyvä ja näkymätön. Nainen Suomen helluntailiikkeen kentällä. - Visible and invisible. Women in the Finnish pentecostal movement. 180 p. Summary 2 p. 2005.
- 263 HEIMONEN, SIRKKALIISA, Työikäisenä Alzheimerin tautiin sairastuneiden ja heidän puolisoitensa kokemukset sairauden alkuvaiheessa. - Experiences of persons with early onset Alzheimer's disease and their spouses in the early stage of the disease. 138 p. Summary 3 p. 2005.
- 264 PIIRÖINEN, HANNU, Epävarmuus, muutos ja ammatilliset jännitteet. Suomalainen sosiaalityö 1990-luvulla sosiaalityöntekijöiden tulkinnoissa. - Uncertainty, change and professional tensions. The Finnish social work in the 1990s in the light of social workers' representations. 207 p. Summary 2 p. 2005.
- 265 MÄKINEN, JARMO, Säätiö ja maakunta. Maakuntarahastojärjestelmän kentät ja verkostot. - Foundation and region: Fields and networks of the system of the regional funds. 235 p. Summary 3 p. 2005.
- 266 PETRELIUS, PÄIVI, Sukupuoli ja subjektius sosiaalityössä. Tulkintoja naistyöntekijöiden muistoista. - Gender and subjectivity in social work - interpreting women workers' memories. 67 p. (175 p.) 2005.
- 267 HOKKANEN, TIINA, Äitinä ja isänä eron jälkeen. Yhteishuoltajavanhemmuus arjen kokemuksena. - As a mother and a father after divorce. Joint custody parenthood as an everyday life experience. 201 p. Summary 8 p. 2005.
- 268 HANNU SIRKKILÄ, Elättäjyyttä vai erotiikkaa. Miten suomalaiset miehet legitimoivat parisuhteensa thaimaalaisen naisen kanssa? - Breadwinner or eroticism. How Finnish men legitimize their partnerships with Thai women. 252 p. Summary 4 p. 2005.
- 269 PENTTINEN, LEENA, Gradupuhetta tutkielmaseminaarissa. - Thesis discourse in an undergraduate research seminar. 176 p. Summary 8 p. 2005.
- 270 KARVONEN, PIIRKKO, Päiväkotilasten lukuleikit. Lukutaidon ja lukemistietoisuuden kehittyminen interventiotutkimuksessa- Reading Games for Children in Daycare Centers. The Development of Reading Ability and Reading Awareness in an Intervention Study . 179 p. Summary 3 p. 2005.
- 271 KOSONEN, PEKKA A., Sosiaalialan ja hoitotyön asiantuntijuuden kehitysehdot ja opiskelijavalinta. - Conditions of expertise development in nursing and and social care, and criteria for student selection. 276 p. Summary 3 p. 2005.

- 272 NIIRANEN-LINKAMA, PÄIVI, Sosiaalisen transformaatio sosiaalialan asiantuntijuu- den diskurssissa. - Transformation of the social in the discourse of social work expertise. 200 p. Summary 3 p. 2005.
- 273 KALLA, OUTI, Characteristics, course and outcome in first-episode psychosis. A cross-cultural comparison of Finnish and Spanish patient groups. - Ensiker- talaisten psykoosipotilaiden psyykkis- sosiaaliset ominaisuudet, sairaudenkulku ja ennuste. Suomalaisten ja espanjalaisten potilasryhmien vertailu. 75 p. (147 p.) Tiivistelmä 4 p. 2005.
- 274 LEHTOMÄKI, ELINA, Pois oppimisyhteiskun- nan marginaalista? Koulutuksen merkitys vuosina 1960-1990 opiskelleiden lapsuu- desta kuurojen ja huonokuuloisten aikuisten elämänculussa. - Out from the margins of the learning society? The meaning of education in the life course of adults who studied during the years 1960-1990 and were deaf or hard-of-hearing from childhood. 151 p. Summary 5 p. 2005.
- 275 KINNUNEN, MARJA-LIISA, Allostatic load in relation to psychosocial stressors and health. - Allostaattinen kuorma ja sen suhde psykososiaalisiin stressitekijöihin ja terveyteen. 59 p. (102 p.) Tiivistelmä 3 p. 2005.
- 276 UOTINEN, VIRPI, I'm as old as I feel. Subjective age in Finnish adults. - Olen sen ikäinen kuin tunnen olevani. Suomalaisten aikuis- ten subjektiivinen ikä. 64 p. (124 p.) Tiivistelmä 3 p. 2005.
- 277 SALOKOSKI, TARJA, Tietokonepelit ja niiden pelaaminen. - Electronic games: content and playing activity. 116 p. Summary 5 p. 2005.
- 278 HIIHNALA, KAUKO, Laskutehtävien suoritta- misesta käsitteiden ymmärtämiseen. Perus- koululaisen matemaattisen ajattelun kehittyminen aritmetiikasta algebraan siirryttäessä. - Transition from the performing of arithmetic tasks to the understanding of concepts. The development of pupils' mathematical thinking when shifting from arithmetic to algebra in comprehensive school. 169 p. Summary 3 p. 2005.
- 279 WALLIN, RISTO, Yhdistyneet kansakunnat organisaationa. Tutkimus käsitteellisestä muutoksesta maailmanjärjestön organi- soinnin periaatteissa. - From the league to UN. The move to an organizational vocabulary of international relations. 172 p. Summary 2 p. 2005.
- 280 VALLEALA, ULLA MAIJA, Yhteinen ymmär- täminen koulutuksessa ja työssä. Kontekstin merkitys ymmärtämisessä opiskelijaryh- män ja työtiimin keskusteluissa. - Shared understanding in education and work. Context of understanding in student group and work team discussions. 236 p. Summary 7 p. 2006.
- 281 RASINEN, TUUJA, Näkökulmia vieraskieliseen perusopetukseen. Koulun kehittämishank- keesta koulun toimintakulttuuriksi. - Perspectives on content and language integrated learning. The impact of a development project on a school's activities. 204. Summary 6 p. 2006.
- 282 VIHOLAINEN, HELENA, Suvussa esiintyvän lukemsvaikeusriskin yhteys motoriseen ja kielelliseen kehitykseen. Tallaako lapsi kielensä päälle? - Early motor and language development in children at risk for familial dyslexia. 50 p. (94 p.) Summary 2 p. 2006.
- 283 KIILLI, JOHANNA, Lasten osallistumisen voimavarat. Tutkimus Ipanoiden osallistu- misesta. - Resources for children's participation. 226 p. Summary 3 p. 2006.
- 284 LEPPÄMÄKI, LAURA, Tekijänoikeuden oikeut- taminen. - The justification of copyright. 125 p. Summary 2 p. 2006.
- 285 SANAKSENAHO, SANNA, Eriarvoisuus ja luottamus 2000-luvun taitteen Suomessa. Bourdieulainen näkökulma. - Inequality and trust in Finland at the turn of the 21st century: Bourdieuan approach. 150 p. Summary 3 p. 2006.
- 286 VALKONEN, LEENA, Millainen on hyvä äiti tai isä? Viides- ja kuudesluokkalaisten lasten vanhemmuuskäsitykset. - What is a good father or good mother like? Fifth and sixth graders' conceptions of parenthood. 126 p. Summary 5 p. 2006.
- 287 MARTIKAINEN, LIISA, Suomalaisten nuorten aikuisten elämään tyytyväisyyden monet kasvot. - The many faces of life satisfaction among Finnish young adults. 141 p. Summary 3 p. 2006.
- 288 HAMARUS, PÄIVI, Koulukiusaaminen ilmiönä. Yläkoulun oppilaiden kokemuksia kiusaamisesta. - School bullying as a phenomenon. Some experiences of Finnish lower secondary school pupils. 265 p. Summary 6 p. 2006.
- 289 LEPPÄNEN, ULLA, Development of literacy in kindergarten and primary school. Tiivistelmä 2 p. 49 p. (145 p.) 2006.
- 290 KORVELA, PAUL-ERIK, The Machiavellian reformation. An essay in political theory. 171 p. Tiivistelmä 2 p. 2006.
- 291 METSOMÄKI, MARJO, "Suu on syömistä varten". Lasten ja aikuisten kohtaamisia ryhmäperhepäiväkodin ruokailutilanteissa. - Encounters between children and adults in group family day care dining situations. 251 p. Summary 3 p. 2006.
- 292 LATVALA, JUHA-MATTI, Digitaalisen kommuni- kaatiosovelluksen kehittäminen kodin ja koulun vuorovaikutuksen edistämiseksi. - Development of a digital communication system to facilitate interaction between home and school. 158 p. Summary 7 p. 2006.

- 293 PITKÄNEN, TUULI, Alcohol drinking behavior and its developmental antecedents. - Alkoholien juomiskäyttäytyminen ja sen ennustaminen. 103 p. (169 p.) Tiivistelmä 6 p. 2006.
- 294 LINNILÄ, MAIJA-LIISA, Kouluvalmiudesta koulun valmiuteen. Poikkeuksellinen koulunaloitus koulumenestyksen, viranomaislausuntojen ja perheiden kokemusten valossa. - From school readiness to readiness of school - Exceptional school starting in the light of school attainment, official report and family experience. 321 p. Summary 3 p. 2006.
- 295 LEINONEN, ANU, Vanhusneuvoston funktioita jäljittämässä. Tutkimus maaseutumaisten kuntien vanhusneuvostoista. - Tracing functions of older people's councils. A study on older people's councils in rural municipalities. 245 p. Summary 3 p. 2006.
- 296 KAUPPINEN, MARKO, Canon vs. charisma. "Maoism" as an ideological construction. - Kaanon vs. karisma. "Maoismi" ideologisenä konstruktiona. 119 p. Yhteenveto 2 p. 2006.
- 297 VEHKAKOSKI, TANJA, Leimattu lapsuus? Vammaisuuden rakentuminen ammatti-ihmistien puheissa ja teksteissä. - Stigmatized childhood? Constructing disability in professional talk and texts. 83 p. (185 p.) Summary 4 p. 2006.
- 298 LEPPÄÄHO, HENRY, Matemaattisen ongelman ratkaisutaidon opettaminen peruskoulussa. Ongelmanratkaisukurssin kehittäminen ja arviointi. - Teaching mathematical problem solving skill in the Finnish comprehensive school. Designing and assessment of a problem solving course. 343 p. Summary 4 p. 2007.
- 299 KUVAJA, KRISTIIINA, Living the Urban Challenge. Sustainable development and social sustainability in two southern megacities. 130 p. (241 p.) Yhteenveto 4 p. 2007.
- 300 POHJOLA, PASI, Technical artefacts. An ontological investigation of technology. 150 p. Yhteenveto 3 p. 2007.
- 301 KAUKUA, JARI, Avicenna on subjectivity. A philosophical study. 161 p. Yhteenveto 3 p. 2007.
- 302 KUPILA, PÄIVI, "Minäkö asiantuntija?". Varhaiskasvatuksen asiantuntijan merkitysperspektiivin ja identiteetin rakentuminen. - "Me, an expert?" Constructing the meaning perspective and identity of an expert in the field of early childhood education. 190 p. Summary 4 p. 2007.
- 303 SILVENNOINEN, PIIA, Ikä, identiteetti ja ohjaava koulutus. Ikääntyvät pitkäaikaistyöttömät oppimisyhteiskunnan haasteena. - Age, identity and career counselling. The ageing, long-term unemployed as a challenge to learning society. 229 p. Summary 4 p. 2007.
- 304 REINIKAINEN, MARJO-RIITTA, Vammaisuuden sukupuolittuneet ja sortavat diskurssit: Yhteiskunnallis-diskursiivinen näkökulma vammaisuuteen. - Gendered and oppressive discourses of disability: Social-discursive perspective on disability. 81 p. (148 p.) Summary 4 p. 2007.
- 305 MÄÄTTÄ, JUKKA, Asepalvelus nuorten naisten ja miesten opinto- ja työuralla. - The impact of military service on the career and study paths of young women and men. 141 p. Summary 4 p. 2007.
- 306 PYYKKÖNEN, MIIKKA, Järjestäytyvät diasporat. Etnisyys, kansalaisuus, integraatio ja hallinta maahanmuuttajien yhdistystoiminnassa. - Organizing diasporas. Ethnicity, citizenship, integration, and government in immigrant associations. 140 p. (279 p.) Summary 2 p. 2007.
- 307 RASKU, MINNA, On the border of east and west. Greek geopolitical narratives. - Idän ja lännen rajalla. Narratiiveja kreikkalaisesta geopolitiikasta. 169 p. Yhteenveto 3 p. 2007.
- 308 LAPIOLAHTI, RAIMO, Koulutuksen arviointi kunnallisen koulutuksen järjestäjän tehtävänä. Paikallisen arvioinnin toteutumisedellytysten arviointia erään kuntaorganisaation näkökulmasta. - The evaluation of schooling as a task of the communal maintainer of schooling - what are the presuppositions of the execution of evaluation in one specific communal organization. 190 p. Summary 7 p. 2007.
- 309 NATALE, KATJA, Parents' Causal Attributions Concerning Their Children's Academic Achievement. - Vanhempien lastensa koulumenestystä koskevat kausaaliattribuutit. 54 p. (154 p.) Yhteenveto 3 p. 2007.
- 310 VAHTERA, SIRPA, Optimistit opintiellä. Opinnoissaan menestyvien nuorten hyvinvointi lukiosta jatko-opintoihin. - The well-being of optimistic, well-performing high school students from high school to university. 111 p. Summary 2 p. 2007.
- 311 KOIVISTO, PÄIVI, "Yksilöllistä huomiota arkisissa tilanteissa". Päiväkodin toimintakulttuurin kehittäminen lasten itsetuntoa vahvistavaksi. - "Individual attention in everyday situations". Developing the operational culture of a day-care centre to strengthen children's self-esteem. 202 p. Summary 4 p. 2007.
- 312 LAHIKAINEN, JOHANNA, "You look delicious" - Food, eating, and hunger in Margaret Atwood's novels. 277 p. Yhteenveto 2 p. 2007.
- 313 LINNAVUORI, HANNARIIKKA, Lasten kokemuksia vuoroasumisesta. - Children's experiences of dual residence. 202 p. Summary 8 p. 2007.
- 314 PARVIAINEN, TIINA, Cortical correlates of language perception. Neuromagnetic studies in adults and children. - Kielen käsittely aivoissa. Neuromagneettisia tutkimuksia aikuisilla ja lapsilla. 128 p. (206 p.) Yhteenveto 5 p. 2007.

- 315 KARA, HANNELE, *Ermutige mich Deutsch zu sprechen. Portfolio als evaluationsform von mündlichen leistung.* - "Rohkaise minua puhumaan saksaa" - kielisalkku suullisen kielitaidon arviointivälineenä. 108 p. Yhteenveto 3 p. 2007.
- 316 MÄKELÄ, AARNE, *Mitä rehtorit todella tekevät. Etnografinen tapaustutkimus johtamisesta ja rehtorin tehtävistä peruskoulussa.* - What principals really do. An ethnographic case study on leadership and on principal's tasks in comprehensive school. 266 p. Summary 5 p. 2007.
- 317 PUOLAKANAHO, ANNE, *Early prediction of reading - Phonological awareness and related language and cognitive skills in children with a familial risk for dyslexia.* - Lukemistaitojen varhainen ennustaminen. Fonologinen tietoisuus, kielelliset ja kognitiiviset taidot lapsilla joiden suvussa esiintyy dysleksiaa. 61 p. (155 p.) Yhteenveto 3 p. 2007.
- 318 HOFFMAN, DAVID M., *The career potential of migrant scholars in Finnish higher education. Emerging perspectives and dynamics.* - Akateemisten siirtolaisten uramahdollisuudet suomalaisessa korkeakoulujärjestelmässä: dynamiikkaa ja uusia näkökulmia. 153 p. (282 p.) Yhteenveto 2 p. 2007.
- 319 FADJUKOFF, PÄIVI, *Identity formation in adulthood.* - Identiteetin muotoutuminen aikuisiässä. 71 p. (168 p.) Yhteenveto 5 p. 2007.
- 320 MÄKIKANGAS, ANNE, *Personality, well-being and job resources: From negative paradigm towards positive psychology.* - Persoonallisuus, hyvinvointi ja työn voimavarat: Kohti positiivista psykologiaa. 66 p. (148 p.) Yhteenveto 3 p. 2007.
- 321 JOKISAARI, MARKKU, *Attainment and reflection: The role of social capital and regrets in developmental regulation.* - Sosiaalisen pääoman ja toteutumattomien tavoitteiden merkitys kehityksen säätelyssä. 61 p. (102 p.) Yhteenveto 2 p. 2007.
- 322 HÄMÄLÄINEN, JARMO, *Processing of sound rise time in children and adults with and without reading problems.* - Äänten nousuaikojen prosessointi lapsilla ja aikuisilla, joilla on dysleksia ja lapsilla ja aikuisilla, joilla ei ole dysleksiaa. 48 p. (95 p.) Tiivistelmä 2 p. 2007.
- 323 KANERVIO, PEKKA, *Crisis and renewal in one Finnish private school.* - Kriisi ja uudistuminen yhdessä suomalaisessa yksityiskoulussa. 217 p. Tiivistelmä 2 p. 2007.
- 324 MÄÄTTÄ, SAMI, *Achievement strategies in adolescence and young adulthood.* - Nuorten ajattelu- ja toimintastrategia. 45 p. (120 p.) Tiivistelmä 3 p. 2007.
- 325 TORPPA MINNA, *Pathways to reading acquisition: Effects of early skills, learning environment and familial risk for dyslexia.* - Yksilöllisiä kehityspolkuja kohti lukemisen taitoa: Varhaisten taitojen, oppimisympäristön ja sukuriskin vaikutukset. 53 p. (135 p.) 2007.
- 326 KANKAINEN, TOMI, *Yhdistykset, instituutiot ja luottamus.* - Voluntary associations, institutions and trust. 158 p. Summary 7 p. 2007.
- 327 PIRNES, ESA, *Merkityksellinen kulttuuri ja kulttuuripolitiikka. Laaja kulttuurin käsite kulttuuripolitiikan perusteluna.* - Meaningful culture and cultural policy. A broad concept of culture as a basis for cultural policy. 294 p. Summary 2 p. 2008.
- 328 NIEMI, PETTERI, *Mieli, maailma ja referenssi. John McDowellin mielenfilosofian ja semantiikan kriittinen tarkastelu ja ontologinen täydennys.* - Mind, world and reference: A critical examination and ontological supplement of John McDowell's philosophy of mind and semantics. 283 p. Summary 4 p. 2008.
- 329 GRANBOM-HERRANEN, LIISA, *Sananlaskut kasvatuspuheessa - perinnettä, kasvatusta, indoktrinaatiota? - Proverbs in pedagogical discourse - tradition, upbringing, indoctrination?* 324 p. Summary 8 p. 2008.
- 330 KYKYRI, VIRPI-LIISA, *Helping clients to help themselves. A discursive perspective to process consulting practices in multi-party settings.* - Autetaan asiakasta auttamaan itse itseään. Diskursiivinen näkökulma prosessikonstultoinnin käytäntöihin ryhmätilanteissa. 75 p. (153 p.) Tiivistelmä 2 p. 2008.
- 331 KIURU, NOONA, *The role of adolescents' peergroups in the school context.* - Nuortentoveriryhmien rooli kouluympäristössä. 77 p. (192 p.) Tiivistelmä 3 p. 2008.
- 332 PARTANEN, TERHI, *Interaction and therapeutic interventions in treatment groups for intimately violent men.* 46 p. (104 p.) Yhteenveto 2 p. 2008.
- 333 RAITTILA, RAIJA, *Retkellä. Lasten ja kaupunkiympäristön kohtaaminen.* - Making a visit. Encounters between children and an urban environment. 179 p. Summary 3 p. 2008.
- 334 SUME, HELENA, *Perheen pyörteinen arki. Sisäkorvaistutetta käyttävän lapsen matka kouluun.* - Turbulent life of the family. Way to school of a child with cochlear implant. 208 p. Summary 6 p. 2008.
- 335 KOTIRANTA, TUUJA, *Aktivoinnin paradoksit.* - The paradoxes of activation. 217 p. Summary 3 p. 2008.
- 336 RUOPPILA, ISTO, HUUHTANEN, PEKKA, SEITSAMO, JORMA AND ILMARINEN, JUHANI, *Age-related changes of the work ability construct and its relation to cognitive functioning in the older worker: A 16-year follow-up study.* 97 p. 2008.
- 337 TIKKANEN, PIRJO, *"Helpompaa ja haus Kempaa kuin luulin". Matematiikka suomalaisten ja unkarilaisten perusopetuksen neljäsluokkalaisten kokemana.* - "Easier and more fun that

- I thought". Mathematics experienced by fourth-graders in Finnish and Hungarian comprehensive schools. 309 p. Summary 3 p. 2008.
- 338 KAUPPINEN, ILKKA, Tiedon omistaminen on valtaa – Globalisoituvan patenttijärjestelmän poliittinen moraalilaitous ja globaali kapitalismi. – *Owning* knowledge is power. Political moral economy of the globalizing patent system and global capitalism. 269 p. Summary 5 p. 2008.
- 339 KUJALA, MARIA, Muukalaisena omassa maassa. Miten kasvaa vuorovaikutuskonflikteissa? – A stranger in one's own land. How to grow in interaction conflicts? 174 p. Summary 7 p. 2008.
- 340 KOPONEN, TUIRE, Calculation and Language: Diagnostic and intervention studies. – Laskutaito ja kieli: Diagnostinen ja kuntoutustutkimus. 49 p. (120 p.) Tiivistelmä 2 p. 2008.
- 341 HAUTALA, PÄIVI-MARIA, Lupa tulla näkyväksi. Kuvataideterapeuttinen toiminta kouluissa. – Permission to be seen. Art therapeutic activities in schools. 202 p. 2008.
- 342 SIPARI, SALLA, Kuntouttava arki lapsen tueksi. Kasvatuksen ja kuntoutuksen yhteistoiminnan rakentuminen asiantuntijoiden keskusteluissa. – Habilitative everyday life to support the child. Construction of the collaboration of education and rehabilitation in experts discussions. 177 p. Summary 4 p. 2008.
- 343 LEHTONEN, PÄIVI HANNELE, Voimauttava video. Asiakslähtöisyyden, myönteisyyden ja videokuvan muodostama työorientaatio perhetyön menetelmänä. – Empowering video. A work orientation formed by client-focus, positivity and video image as a method for family work. 257 p. Summary 3 p. 2008.
- 344 RUOHOMÄKI, JYRKI, "Could Do Better". Academic Interventions in Northern Ireland Unionism. – "Could Do Better" Akateemiset interventiot Pohjois-Irlannin unionismiin. 238 p. Tiivistelmä 2 p. 2008.
- 345 SALMI, PAULA, Nimeäminen ja lukemisvaikeus. Kehityksen ja kuntoutuksen näkökulma. – Naming and dyslexia: Developmental and training perspectives. 169 p. Summary 2 p. 2008.
- 346 RANTANEN, JOHANNA, Work-family interface and psychological well-being: A personality and longitudinal perspective. – Työn ja perheen vuorovaikutuksen yhteys psyykkiseen hyvinvointiin sekä persoonallisuuteen pitkittäistutkimuksen näkökulmasta 86 p. (146 p.) Yhteenveto 6 p. 2008.
- 347 PIIPPO, JUKKA, Trust, Autonomy and Safety at Integrated Network- and Family-oriented mode for co-operation. A Qualitative Study. 70 p. (100 p.) Yhteenveto 2 p. 2008.
- 348 HÄTINEN, MARJA, Treating job burnout in employee rehabilitation: Changes in symptoms, antecedents, and consequences. – Työuupumuksen hoito työikäisten kuntoutuksessa: muutokset työuupumuksen oireissa, ennakoijissa ja seurauksissa. 85 p. (152 p.) Tiivistelmä 4 p. 2008.
- 349 PRICE, GAVIN, Numerical magnitude representation in developmental dyscalculia: Behavioural and brain imaging studies. 139 p. 2008.
- 350 RAUTIAINEN, MATTI, Keiden koulu? Aineenopettajaksi opiskelevien käsityksiä koulu-kulttuuriin yhteisöllisyydestä. – Who does school belong to? Subject teacher students' conceptions of community in school culture. 180 p. Summary 4 p. 2008.
- 351 UOTINEN, SANNA, Vanhempien ja lasten toimijuuteen konduktiivisessa kasvatuksessa. – Into the agency of a parent and a child in conductive education. 192 p. Summary 3 p. 2008.
- 352 AHONEN, HELENA, Rehtoreiden kertoma johtajuus ja johtajaidentiteetti. – Leadership and leader identity as narrated by headmasters. 193 p. 2008.
- 353 MOISIO, OLLI-PEKKA, Essays on radical educational philosophy. 151 p. Tiivistelmä 3 p. 2009.
- 354 LINDQVIST, RAIIJA, Parisuhdeväkivallan kohtaaminen maaseudun sosiaalityössä. – Encountering partner violence with rural social work. 256 p. 2009.
- 355 TAMMELIN, MIA, Working time and family time. Experiences of the work and family interface among dual-earning couples in Finland. – Työaika ja perheen aika: kokemuksia työn ja perheen yhteensovittamisesta Suomessa. 159 p. Tiivistelmä 3 p. 2009.
- 356 RINNE, PÄIVI, Matkalla muutokseen. Sosiaalialan projektitoiminnan perustelut, tavoitteet ja toimintatavat Sosiaaliturva-lehden kirjoituksissa 1990-luvulla. – On the way to the change. 221 p. Summary 2 p. 2009.
- 357 VALTONEN, RIITTA, Kehityksen ja oppimisen ongelmien varhainen tunnistaminen Lene-arvioinnin avulla. Kehityksen ongelmien päällekkäisyys ja jatkuvuus 4–6-vuotiailla sekä ongelmien yhteys koulusuoriutumiseen. – Lene-assessment and early identification of developmental and learning problems. Co-occurrence and continuity of developmental problems from age 4 to age 6 and relation to school performance. 73 p. (107 p.) Summary 2 p. 2009.
- 358 SUHONEN, KATRI, Mitä hiljainen tieto on hengellisessä työssä? Kokemuksellinen näkökulma hiljaisen tiedon ilmenemiseen, siirrettävyyteen ja siirrettävyyden merkitykseen ikääntyneiden diakoniatyöntekijöiden ja pappien työssä. – What is tacit knowledge in spiritual work? An experiential approach to the manifestation, significance and distribution of tacit knowledge in the work of aged church deacons and ministers. 181 p. Summary 6 p. 2009.

- 359 JUMPPANEN, AAPO, United with the United States - George Bush's foreign policy towards Europe 1989-1993. 177 p. Yhteenvedo 3 p. 2009.
- 360 HUEMER, SINI, Training reading skills. Towards fluency. - Lukemistaitojen harjoittaminen. Tavoitteena sujuvuus. 85 p. (188 p.) Yhteenvedo 3 p. 2009.
- 361 ESKELINEN, TEPPO, Putting global poverty in context. A philosophical essay on power, justice and economy. 221 p. Yhteenvedo 1 p. 2009.
- 362 TAIPALE, SAKARI, Transformative technologies, spatial changes: Essays on mobile phones and the internet. 97 p. (184 p.) Yhteenvedo 3 p. 2009.
- 363 KORKALAINEN, PAULA, Riittämättömyyden tunteesta osaamisen oivallukseen. Ammatillisen asiantuntijuuden kehittäminen varhaiserityiskasvatuksen toimintaympäristöissä. - From a feeling of insufficiency to a new sense of expertise. Developing professional knowledge and skills in the operational environments for special needs childhood education and care. 303 p. Summary 4 p. 2009.
- 364 SEPPÄLÄ-PÄNKÄLÄINEN, TARJA, Oppijoiden moninaisuuden kohtaaminen suomalaisessa lähikoulussa. Etnografia kouluyhteisön aikuisten yhdessä oppimisen haasteista ja mahdollisuuksista. - Confronting the Diversity of Learners in a Finnish Neighbourhood School. An Ethnographic Study of the Challenges and Opportunities of Adults Learning Together in a School community. 256 p. Summary 4 p. 2009.
- 365 SEVÓN, EIJA, Maternal Responsibility and Changing Relationality at the Beginning of Motherhood. - Äidin vastuu ja muuttuvat perhesuhteet äitiyden alussa. 117 p. (200 p.) Yhteenvedo 5 p. 2009.
- 366 HUTTUNEN-SCOTT, TIINA, Auditory duration discrimination in children with reading disorder, attention deficit or both. - Kuulonvarainen keston erottelu lapsilla, joilla on lukemisvaikeus, tarkkaavaisuuden ongelma tai molemmat. 68 p. (112 p.) Tiivistelmä 3 p. 2009.
- 367 NEUVONEN-RAUHALA, MARJA-LIISA, Työelämälähtöisyyden määrittäminen ja käyttäminen ammattikorkeakoulun jatkotutkintokokeilussa. - Defining and applying working-life orientation in the polytechnic postgraduate experiment. 163 p. Summary 7 p. 2009.
- 368 NYMAN, TARJA, Nuoren vieraan kielen opettajan pedagogisen ajattelun ja ammatillisen asiantuntijuuden kehittyminen. - The development of pedagogical thinking and professional expertise of newly qualified language teachers. 121 p. (201 p.) Summary 4 p. 2009.
- 369 PUUTIO, RISTO, Hidden agendas. Situational tasks, discursive strategies and institutional practices in process consultation. 83 p. (147 p.) Tiivistelmä 2 p. 2009.
- 370 TOIVANEN, JUHANA, Animal consciousness. Peter Olivi on cognitive functions of the sensitive soul. 369 p. Yhteenvedo 4 p. 2009.
- 371 NOKIA, MIRIAM, The role of the hippocampal theta activity in classical eyeblink conditioning in rabbits. - Hippokampuksen theta-aktiivisuuden rooli klassisessa silmäniskuehdollistamisessa kaneilla. 41 p. (80 p.) Yhteenvedo 2 p. 2009.
- 372 LÄHTEENMÄKI, VILI, Essays on early modern conceptions of consciousness: Descartes, Cudworth, and Locke. 160 p. 2009.
- 373 BJÖRK, KAJ, What explains development. Development strategy for low human development index countries. 212 p. Yhteenvedo 1 p. 2009.
- 374 PUUPPONEN, ANTTI, Maaseutuyrittäjyys, verkostot ja paikallisuus. Tapaustutkimus pienimuotoisen elintarviketuotannon kestävyyydestä Keski-Suomessa. - Rural entrepreneurship, networks and locality. A case study of the sustainability of small-scale food production in Central Finland. 100 p. (191 p.) Summary 3 p. 2009.
- 375 HALTTUNEN, LEENA, Päivähoitotyö ja johtajuus hajautetussa organisaatiossa. - Day care work and leadership in a distributed organization. 181 p. Summary 4 p. 2009.
- 376 KAIDESOJA, TUUKKA, Studies on ontological and methodological foundations of critical realism in the social sciences. 65 p. (187 p.) Yhteenvedo 9 p. 2009.
- 377 SIPPOLA, MARKKU, A low road to investment and labour management? The labour process at Nordic subsidiaries in the Baltic States. 272 p. Tiivistelmä 2 p. 2009.
- 378 SANTALA, OLLI-PEKKA, Expertise in using the Rorschach comprehensive system in personality assessment. 150 p. Tiivistelmä 1 p. 2009.
- 379 HARJUNEN, HANNELE, Women and fat: Approaches to the social study of fatness. - Naiset ja lihavuus: näkökulmia lihavuuden yhteiskuntatieteelliseen tutkimukseen 87 p. (419 p.) Tiivistelmä 4 p. 2009.
- 380 KETTUNEN, LIISA, Kyllä vai ei. Peruskoulun sukupuolikasvatuksen oppimateriaalin kehittämistyö ja arviointi. - Yes or no? The development and evaluation of teaching material for sex education in the Finnish comprehensive school. 266 p. Summary 3 p. 2010.
- 381 FROM, KRISTINE, "Että sais olla lapsena toisten lasten joukossa". Substantiivinen teoria erityistä tukea tarvitsevan lapsen toiminnallisesta osallistumisesta toimintaympäristöissä. - To be a child just as the others in the peer group. A substantive theory of activity-

- based participation of the child with special educational needs. 174 p. Summary 4 p. 2010.
- 382 MYKKÄNEN, JOHANNA, Isäksi tulon tarinat, tunteet ja toimijuus. - Becoming a father – types of narrative, emotions and agency. 166 p. Summary 5 p. 2010.
- 383 RAASUMAA, VESA, Perusopetuksen rehtori opettajien osaamisen johtajana. - Knowledge management functions of a principal in basic education. 349 p. Summary 5 p. 2010.
- 384 SISIÄINEN, LAURI, Foucault's voices: Toward the political genealogy of the auditory-sonorous. - Foucault'n äänet. Kohti auditoris-sonoorista poliittista genealogiaa. 207 p. Tiivistelmä 2 p. 2010.
- 385 PULLI, TUULA, Totta ja unta. Draama puhe- ja kehitysvammaisten ihmisten yhteisöllisenä kuntoutuksena ja kokemuksena. - The Real and the Illusory. Drama as a means of community-based rehabilitation and experience for persons with severe learning and speech disabilities. 281 p. Summary 7 p. 2010.
- 386 SISKONEN, TIINA, Kielelliset erityisvaikeudet ja lukemaan oppiminen. - Specific language impairments and learning to read. 205 p. Summary 3 p. 2010.
- 387 LYYRA, PESSI, Higher-order theories of consciousness: An appraisal and application. - Korkeamman kertaluvun tietoisuusteoria: arvio ja käyttöehdotus. 163 p. Yhteenveto 5 p. 2010.