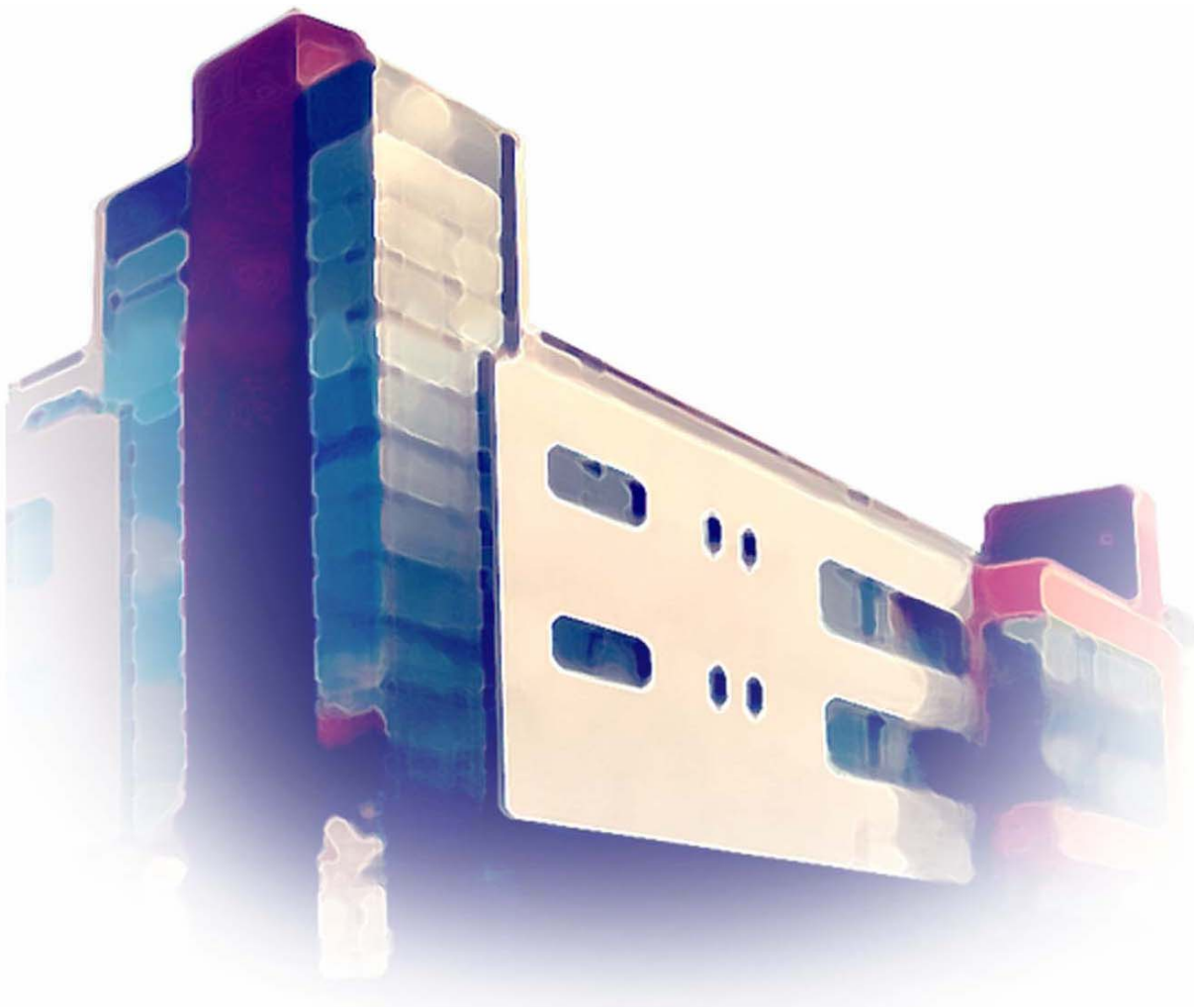


Radoslaw Martin Cichy

*The Dennettian Concept of Intentionality:
Past and Present*



PICS

Publications of the Institute of Cognitive Science

Volume 8-2005

ISSN: 1610-5389

Series title: PICS
Publications of the Institute of Cognitive Science

Volume: 8-2005

Place of publication: Osnabrück, Germany

Date: August 2005

Editors: Kai-Uwe Kühnberger
Peter König
Petra Ludewig

Cover design: Thorsten Hinrichs

The Dennettian Concept of Intentionality: Past and Present

Bachelor Thesis in Cognitive Science

Radoslaw Martin Cichy
University of Osnabrück, Germany

First Supervisor: Prof. Dr. Achim Stephan
Second Supervisor: Dr. Frank Esken

September 12th, 2003

Abstract

The phenomenon of intentionality, i.e. aboutness or of-ness is one of the major topics in contemporary philosophy of mind. In D.C. Dennett's highly influential, stimulating and in its richness unique philosophy the concept of intentionality takes centre stage as a presupposition for any other kind of philosophy of mind.

This thesis critically evaluates the different forms that the Dennettian concept of intentionality has taken over time. It starts discussing Dennett's early instrumentalist philosophy and ends up with his concept of tools and memes. Conceptual relationships *between* subsequent phases in Dennett's development are stressed and given primary role.

In Appreciation of Mr T.W.

| | | |
|-------------|---|-----------|
| 1. | <i>Introduction</i> | 6 |
| 2. | <i>The Early Dennett: Brute Instrumentalism</i> | 13 |
| 2.1. | Origins and Motivation | 13 |
| 2.1.1. | The Intentional Idiom | 13 |
| 2.1.2. | Ryle's Influences | 14 |
| 2.1.3. | Quine's Influences | 14 |
| 2.1.4. | Sellars and the Normative Dimension | 15 |
| 2.1.5. | Predictability and Opposition to Cognitive Psychology | 15 |
| 2.1.6. | The Introduction of Rationality | 16 |
| 2.1.7. | Connections and First Conclusions | 16 |
| 2.2. | Intentional Systems | 17 |
| 2.3. | Arising Problems of the Early Approach | 19 |
| 2.3.1. | General Scepticism Regarding Instrumentalism | 19 |
| 2.3.2. | Criticism Regarding Explicativeness in a Proto-Scientific Theory | 20 |
| 2.3.3. | Criticism Regarding the Moral Project | 22 |
| 2.3.4. | The Accusation of Irrealism | 22 |
| 2.3.5. | Conclusions | 23 |
| 3. | <i>The Later Dennett: Struggle with Objectivity</i> | 24 |
| 3.1. | Objective Interpretation and the Pattern Theory | 24 |
| 3.1.1. | Superiority of One Attribution | 24 |
| 3.1.2. | Are Two Equally Good Interpretations Possible? | 27 |
| 3.2. | The reality of the patterns | 28 |
| 3.3. | Wrap-up | 30 |
| 4. | <i>The Later Dennett: The Relation between the Stances</i> | 30 |
| 4.1. | Intentional system theory | 31 |
| 4.2. | Subpersonal Cognitive Psychology and Its Relation to Intentional System Theory | 32 |
| 4.3. | Criticism | 35 |
| 4.3.1. | 'Too Great a Leap' | 35 |
| 4.3.2. | The Problem of Rationalization | 36 |
| 5. | <i>Rationality and the Evolutionary Turn</i> | 38 |
| 5.1. | Introduction, Early Position and Criticism | 38 |
| 5.2. | Dennett's "Answer" | 41 |
| 5.2.1. | Rationality as Applied to Single Actions | 42 |
| 5.2.2. | Rationality as applied to a System's Design | 42 |
| 5.3. | Wrap-up | 44 |
| 6. | <i>Micro- and Macro-Reduction</i> | 46 |
| 6.1. | Dennett and His Place in the Philosophical Landscape | 46 |
| 6.2. | Fodor's Criticism of Dennett and Dennett's Criticism of Fodor | 48 |
| 6.2.1. | Fodor's Position | 48 |
| 6.2.2. | Dennett, Functionalism and a Critique of Fodor | 49 |
| 6.2.3. | Realism and Causation | 50 |
| 6.2.4. | Wrap-up | 53 |
| 6.3. | Macro-reduction | 53 |
| 7. | <i>Representation, Language and Intentionality Revisited</i> | 56 |
| 7.1. | Introduction | 56 |
| 7.2. | The Genesis of Intentionality and the Conflation of the Stances | 58 |

| | | |
|-------------|---|-----------|
| 7.3. | Language and Representation | 64 |
| 7.3.1. | The Beginnings of Representation | 64 |
| 7.3.2. | Styles of Mental Representation | 66 |
| 7.3.3. | Explicit Representation and Language | 69 |
| 7.4. | Wrap-up | 72 |
| 8. | <i>Tools, Memes and Mind-tools</i> | 73 |
| 8.1. | Introduction | 73 |
| 8.2. | Tools, Mind Tools and Memes | 74 |
| 8.2.1. | Genesis of Language and Deliberate Use | 77 |
| 8.2.2. | Objectification of Content | 79 |
| 8.3. | A ‘Kind of Mind’ | 84 |
| 8.4. | Wrap-up | 86 |
| 9. | <i>A Few Final Words</i> | 87 |
| 10. | <i>References</i> | 88 |

1. Introduction

Daniel C. Dennett has been one of the major philosophers in the field of philosophy of mind over the last decades. His exceptional style of philosophical engagement, his ideas and personality have caught attention in a broad philosophical community, often given new impetus to and shaped philosophical discussion. Though Dennett is an analytic philosopher beyond reasonable doubt, his unique style of philosophical engagement deserves treatment that in my opinion *cannot* be achieved by means of the usual analytical analysis. My aim therefore *is not* to put forward a single hypothesis concerning one of the many philosophical topics that Dennett has discussed over the years, to analyze it exhaustively and to lead it to a culminating proposal of solution or best interpretation. My aim *is* to provide the reader with a different kind of approach in order to do justice to the richness of Dennett's philosophy.

To introduce the reader to the structure and content of this bachelor thesis I offer the following argumentation. Dennett's philosophy has undoubtedly developed strongly over the years. Reading an early and a late text from Dennett one may therefore wonder how these two fit together. Furthermore, since Dennett is a major philosopher his writings have received ample criticism and commentary. Reading these one may wonder whether the criticism posed really fits, whether it hits the core of Dennett's writings or whether it remains superficial. Asking this kind of questions the reader can help himself to further literature from an immense body, both of primary kind written by Dennett himself as well as of secondary kind written by his critics. Doing so the reader will probably make the initial experience of being lost in a huge welter of unconnected, fragmentary and in many aspects unsatisfying technical writings which in some way seem to fit together, but in others do not. In this way being lost the reader may ask himself why in the end he shall be persuaded by Dennett and not by his critics, or vice versa.

This bachelor thesis has been written out of the experience of this loss of orientation and is meant to provide a discussion of Dennett's work that is able to amend this situation. Just a further purely analytical and fragmentary text I feel will obviously not be able to amend the situation, but would more probably increase the existing confusion. For that reason the reader shall not expect a single answer to a question regarding a typically 'Dennettian' problem, neither a loose collection of answers to the questions surrounding Dennett's philosophy.

What the reader may expect is a kind of text that wants to do justice to Dennett's philosophy in several aspects. First, it has the aim to take into account Dennett's unique style of philosophical engagement in its discussion instead of providing the reader exclusively with

a discussion from some or other philosophical point of view. Second, it has the aim to show Dennett's position as compared to other major philosophers of mind in order to locate it in a philosophical landscape. Third, it wants to show that this kind of engagement is more fruitful than any other in understanding the richness and complexity of Dennett's philosophy.

The basis for all of Dennett's writings undoubtedly is his treatment of the concept of *intentionality*. It is from here that any approach towards an understanding and thorough analysis of Dennettian philosophy has to take a start. Dennett's writings on consciousness and morality as well as his engagement in the philosophy of AI presuppose his prior concept of intentionality and are couched in terms connected to it. This is already visible in the title of his first book 'Content and Consciousness' (Dennett 1969) – it is *first* content, *then* consciousness that has to be explained. Hence, before any analysis of Dennett's concept of consciousness can be carried out, his philosophy of intentionality has to be considered. *This thesis therefore makes an attempt to get clear on Dennett's concept of intentionality both in its past, as well as in its present forms.*

Such a project faces serious difficulties, both technically as well as in its endeavour to convince. Dennett being a strongly systematic philosopher in the sense of construing strong interrelationships and bonds between the concepts involved, it seems that one cannot explain Dennettian intentionality without in some sense also explaining the rest of Dennett, too. To avoid such a presumptuous claim I have adopted a policy towards the project which I believe to be best made clear by an allegory between Dennett and his philosophy and a chess player and the game of chess.

Besides being an engaged and widely acknowledged philosopher of mind Dennett also is a 'player' and the form of his philosophical writings are a clear account of the fact that to him philosophizing is 'fun'. It is not a struggle against insurmountable odds, it is not a distressing activity. Dennett's style of writing is vivid, sometimes humorous, easy and fluid. I propose that viewing Dennett as a player opens up possibilities to look at his philosophy in a different than usual and more interesting way.

As a good chess player, Dennett has learned chess from the masters. The topics he discusses, the style in which he writes are a mirror of the philosophical air of the time. Chapter 2.1 is meant to furnish the reader with the most important influences that are visible in Dennett's work. This localization of Dennett in terms of his major teachers and influences is helpful in creating a basic answer to the question of *why* Dennett does write the way he does and discusses the topics he does discuss. The professional philosopher well acquainted with Dennett may want to skip this part, but to other readers it might provide a helpful

orientation. The role these influences play in Dennett's philosophy will be pointed out repeatedly in this thesis, since I believe that a big part of Dennett's philosophy is best understood as being motivated by exactly these influences.

Having received a classical and rigid training is not enough to be a good player. The player has to possess a creative and inventive mind, too, i.e. he has to have 'good ideas' at the right time. Dennett is not only a heavy worker, but also a believer in the power of good ideas. His probably most influential essay is his early 'Intentional Systems' (Dennett 1971) and might be considered to be a 'good idea'. Already in this essay the young Dennett shows his unique style, his fascinating ability to combine the insights of several philosophers into a new and forceful position. Since intentionality is the general topic, and 'Intentional Systems' Dennett's first groundbreaking paper, it will be dealt with early in chapter 2.2 of this thesis.

As any good player, Dennett learns from his mistakes and criticism- one should not make the same mistake twice if one wants to win the game. Naturally Dennett's philosophy has been subject to criticism throughout its development. Dennett himself has felt the force of this criticism and adjusted his philosophy accordingly.

At this point I feel it to be necessary to provide the reader with further information that motivates the structure of this thesis. Reading and evaluating any of the Dennettian texts that are being discussed in this thesis one may, by allegory, ask oneself why Dennett does make such or such a move in his argumentation? One method to find an answer is to take into account only a small piece in the long developmental span of Dennett's writings and ignore the past. In this way one may on the one hand gain forceful insights, but on the other also *lose a very important part in the understanding of Dennett*. Historical one-dimensionality as compared to a diachronic approach sometimes is often not very helpful. It is simply too narrow. Repeatedly during this thesis I will argue that in many cases well known criticism of Dennett's philosophy arises out of this narrow context and can be shown to be superficial if looked at in a *greater context*. Though Dennett's 'moves' thereby will be described in detail, it is not in the manner of *l'art pour l'art*, but first to convey to the reader a broader context in which I find it more satisfying to read Dennett by avoiding narrowness and second to make clear the 'beauty' and 'elegance' of the style in which Dennett plays. Consequently, moves that a player makes and that appear to be weak can be the strongest ones possible if one knows the prior setting of the game and the tactics that the player has in mind. Such knowledge may make the engagement with the game more fruitful and interesting. I believe this to be so in Dennett's case and therefore put forward a diachronic approach in which exposition, criticism and renewed exposition take turns.

A good chess player does not consider the figures on the board one by one, but keeps an eye on the whole board. To him the figures are not fragmented pieces, but comprise a holistic and global whole in which he acts. A move in a game does not only change the position of a figure, but possibly changes the relations between all the figures on the board. A good chess player is sensitive to these changes in relations. The same is true for Dennett. Dennett writes on several topics in philosophy, but in contrast to many other modern philosophers he does so in a highly systematic, complex and interwoven fashion. From this reason studying Dennett one faces a difficulty and an enormous possibility at once. On the one hand it is hard to get a clear picture of Dennett, since everything relates to everything else and therefore many contradictions and questions emerge. On the other hand it is exactly Dennett's avoidance of one-dimensionality and narrowness that makes him such a fascinating philosopher. In a philosophical context that is coined by fragmentary style and unsatisfying technical commentary Dennett's philosophy offers a refreshing and captivating opportunity to gain philosophical insight.

If one sets out to describe the work and style of a player or philosopher of such kind and does not want to re-invent the wheel, it is helpful to draw comparisons to the other major well known players. A player's overall development is mainly shaped by the experiences he makes playing chess with other players. The same can be said about Dennett. To understand Dennett's way of argumentation as well as his development it is necessary to understand how criticism shaped Dennett's philosophy as an answer to this criticism. Main critics of Dennett are introduced in this thesis and it is helpful to see in which respects Dennett *differs* from his critics and in which ways he is *in accord with* them. This thesis makes use of this tool repeatedly.

Given the argumentation above, my aim is to provide the reader with a *critical commentary review* on the philosopher Dennett. In doing so I have made the greatest effort not to express myself in more words than I feel necessary so as to enable the reader to enjoy the full richness of Dennett's philosophy. Nevertheless, the reader who is expecting a discussion that will bring the thesis to a check mate will be disappointed. In other words, the aim of this thesis is to invite the reader to share a critical fascination that Dennett's philosophy can provoke. In this sense this thesis does not have a culmination, no final point that has to be reached, but it is a *journey* through Dennettian philosophy, full of good ideas, critical elucidation, proposals for best interpretation, motivated and well founded which the reader is invited to join. It is supposed to be the kind of text which I often painfully missed

studying philosophy, a text that ‘brings the major pieces together’ and gives insight into the connections and interrelations of a philosophical work.

Being written in such a spirit this text is not aimed at complete beginners in the study of Dennett’s philosophy. Basic knowledge of Dennettian philosophy as well as knowledge of the main positions, problems and philosophers in the field of philosophy of mind is required. Though knowledge of the main questions surrounding the debate of Dennettian philosophy is helpful, it is not mandatory, since the positions of major critics are introduced if necessary. Besides, minor points which do not warrant a detailed exposition are exhaustively furnished with references which the interested reader may want to look up if interested.

Having provided the reader with an exposition of the general spirit of this thesis I would like to return to the exposition of its content. As already mentioned, Dennett’s probably most important essay has been ‘Intentional Systems’ (Dennett 1971), introducing the concept of intentionality as a pragmatic and instrumentalistic tool for behavioural prediction. It is set up in chapter 2.2.

Chapter 2.3 in turn discusses different sorts of criticism that has been formulated against Dennett’s early theory. I will argue that though general criticism regarding instrumentalism can be rebutted in Dennett’s case (Chapter 2.3.1), criticism regarding the explicativeness of Dennett theory (Chapter 2.3.2), the project to allow for a basis for moral ascription (Chapter 2.3.3) and the reality of the ascribed states (Chapter 2.3.4) were felt to have force. I propose that Dennett’s subsequent theory is best understood as an answer to this criticism. This answer is two fold.

First, it comprises a *positive* part, i.e. Dennett’s making his own theory plausible, and a *negative* part, i.e. Dennett’s showing all other approaches to be implausible. The positive part is dealt with in Chapter 3 and 4 which cover Dennett’s pattern theory and his conceptualization of sub-personal cognitive psychology including the *illata/ abstracta* distinction respectively. Chapter 3 answers questions regarding the ‘reality’ status of intentional states in Dennett theory, whereas Chapter 4 deals with questions regarding the relations between the stances and causal efficacy.

In debating these issues I will recognize criticism formulated by Yu and Fuller (Yu and Fuller 1986) to fail to rebut Dennett. Nevertheless, Yu and Fuller recognize a major problem in Dennett’s *abstracta/illata* distinction when it comes to the concept of normativity and rationalization. Treatment of these issues makes a discussion of Dennett’s concept of rationality mandatory which fills Chapter 5.

Dennett's concept of rationality invites confusion between the notions of rationality of actions and rationality of design. To clear up the confusion I put forward the thesis that Dennett is best interpreted as having a concept of rationality of design. Elton's (Elton 2003) canonical suggestion to interpret Dennett's philosophy in terms of rationality of actions will be rebutted.

As already hinted at, Dennett's task in making his position convincing is twofold: it comprises a positive as well as a negative part. Chapter 6.1-6.2 is devoted to the negative part. Chapter 6.1 positions Dennett in a philosophical landscape in general to make clear the major differences between his approach and other approaches on the 'philosophical market'. It is necessary to get a clear understanding of Dennett's exact position in order to be able to appreciate the necessity of the complexity of Dennett's further philosophical development and to understand the major differences between Dennett's position and functionalist argumentations for multiple realization, e.g. as put forward by Putnam or Fodor. Due to lack of scrutiny these very different approaches have often been confused. In order to make clear the connections and crucial differences between Dennett's position and functionalism Fodor's position is used as a foil. Fodor has been a perennial critic of Dennett's position and Dennett has often formulated his views in open opposition to Fodor.

This discussion focuses mainly on Dennett's concept of causality in order to rebut the canonical reading of Dennett's position as allowing for a free flipping between the stances. A more rigid connection between the stances is proposed instead. Furthermore, in chapter 6.3 Dennett's position with regard to alternative approaches to intentionality is discussed in such a way as to show Dennett's motivation to turn to the concepts of macro-reduction and evolutionary psychology. I plan to show these motivations in turn to be the underlying motor of Dennett's newest philosophy and therefore understanding of the context of their genesis to be crucial in understanding Dennettian philosophy.

In this context Dennett develops a theory of language and representation in order to explain the kind of intentionality involved here. These theories are tackled in chapter 7 and 8. They form a highly complex and interrelated system and from my point of view comprise the best exposition of the interrelatedness of Dennett's philosophical system in this thesis. This high complexity and richness is not a sign of conceptual confusion, but emerges naturally as a product of the rigid limitations that Dennett puts on his own work. Though some of the parts may appear unimportant to the reader, I have put greatest effort in the task to make clear in which way this interrelatedness of Dennett's work makes this kind of treatment necessary. I believe that any kind of simpler treatment of the topic would not do justice to Dennett and be

an open, question-begging invitation to criticism on behalf of overt inconsistency due to purportedly missing links in the argumentation. Not the single ideas in Dennett's newer writings are most important, since these are commonly known and held by many philosophers, but it is the way in which Dennett relates these ideas to one another in an enormous system which deserves attention.

Rebutting any other approach to the explanation of intentionality, Dennett puts a great burden upon his shoulders. He is in need of providing a unified and stringent approach to intentionality that both explains *similarities* with regard to intentionality across creatures *and* differences with the same regard. This problem is recognized in chapter 7.1 and the unique nature of human intentionality as compared to all other creatures is emphasised. Chapter 7.2 introduces to the reader Dennett's concept of genesis of intentionality. In debating this issue I will put forward the thesis that Dennett's concept of intentionality finally completely changes into a *teleological* concept. This change will be recognized to be responsible for the form of evolutionary explanation that Dennett's newest philosophy has taken. In discussing Dennett's account of the genesis of intentionality I will show the concept of representation to play a crucial role. Chapter 7.3 will be devoted to the notion of representation and its role in Dennett's account.

Chapter 8 is a continuation of Dennett's project of providing for a theory that is able to explain (human) intentionality by naturalistic means. The concept of tools and the concept of memes as major parts of this project will therefore be introduced (Chapter 8.2). On basis of this introduction I will tackle major questions that emerged during discussion of Dennett's theory in chapter 7. I will argue that Dennett's answer to these questions is highly fascinating, but unfortunately partly unsuccessful (Chapter 8.2-8.2.2). To close the discussion I will restate the question in how far Dennett's newest writings concerning human intentionality are convincing by examining the relation between Dennett's concept of tools and the human mind.

To sum up, I hope to provide the reader with a thesis comprising a broad and in-depth discussion and criticism of Dennettian philosophy of intentionality which differs from the canonical reception of Dennett's work in many important ways. It can be looked on as a first and basic part of a larger thesis whose second and third part would deal with Dennett's concept of consciousness, artificial intelligence and an analysis of the special take Dennett has on adaptationism. Sorrowfully such a project forbids itself due to limitations of length and time. Nevertheless, I am glad to present to the reader this thesis in its present form.

2. The Early Dennett: Brute Instrumentalism

2.1. Origins and Motivation

Dennett, as any philosopher, is a child of his time. Hence, it is necessary to understand in which context Dennett was introduced to the concept of intentionality and how his initial understanding thereof was. I propose to view the context in which Dennett developed his theory to be guided by the following considerations.

2.1.1. **The Intentional Idiom**

Dennett's interest in intentionality is shaped by the idea that intentionality in the first place is mainly a feature of sentences, not yet mental states. The task is therefore considered to be to explain the relation between the truths expressed in mental language encompassing terms like belief, desire, knowledge and thinking, to truths explained in scientific, non-intentional language. The beginning of the discussion can be traced back to Chisholm (Chisholm 1956) and his observation that the irreducibility of intentional idioms to non-intentional language could have significance, though Chisholm in unshaken belief in naturalism deduced the irreducibility thesis from this reason surprisingly to be plainly wrong. Dennett on the other hand is convinced of the truth of sentences containing mental vocabulary and cannot follow Chisholm in his outright rebuttal of the existence of intentionality, hence raising question about the already mentioned relation between mental and physical phenomena. What is, then, the relation between saying that a person believes that p and his believing that p?

The discussion of mental sentences in the 60s and 70s, as can be observed historically, tried to answer this question by moving from considering intentionality primarily as a feature of sentences to features of mental states, stating that the sentence "X believes that p" is true if there is a fact to the matter about X that makes the sentence true. It is proposed, in various guises, that for X to believe that p is for X to be in the mental state of believing that p. This suggestion opens up the way towards identity theories, may it be token or type identity theories and functionalist theories as Turing equivalence functionalism (Dennett 1978b: xi-xxii). Dennett, though, refuses to acknowledge the plausibility of any of these positions outright both by direct criticism of the theories as incoherent, in the case of token and type identity theories mirroring Putnam's multiple realization argument and in the case of functionalism similar considerations (Chapter 6.1-6.2). Dennett's unwillingness to accept any

such theory, though, is grounded in convictions that deserve attention as further points, bringing the discussion to the second point.

2.1.2. Ryle's Influences

Dennett is, as one of his scholars, deeply influenced by the philosophy of mind of Gilbert Ryle (Ryle 1949). This is mirrored both in his understanding of philosophy of mind as explicating the most general questions about the mind, i.e. if there are any, and if this is the case, what they could be, and to furnish the foundation and justification of the fundamental metaphysical assumptions that scientific theories unavoidably make. Acknowledging that the mind has something to do with the brain, and since it is the natural sciences whose subject the brain must be, Dennett asks for what can philosophically be regarded as correct when it comes to the nature and character of the mind that is to be investigated in the natural sciences.

The most obvious influences of Ryle upon Dennett hereby are twofold: first the rejection of both metaphysical and methodological Cartesianism, and second the plausibility of Ryle's investigation of the mind in the form of ordinary language philosophy and logical behaviourism. It is the everyday concepts that we have to start with when doing philosophy of mind and the ascription of e.g. beliefs or desires is not to be seen as warranted by the agent in question possessing some intrinsic quality or other, but by external criteria, i.e. the agent's overt behaviour in context. Furthermore highly influenced by Ryle's observation that causal explanations of phenomena do not coincide with naming a cause in the conceptual analysis, Dennett denies the outright plausibility of causal explanations of intentionality in terms of causal interactions of some kind as e.g. prominently at this time put forward by Fodor in the form of the asymmetric dependency thesis (Fodor 1987).

The differences to Ryle are also apparent. First, Dennett's holistic approach to identification of mental phenomena contradicts Ryle's endeavour to individuate every single mental state behaviouristically in isolation. Second, Dennett's bigger openness to the relevance of science in the explanation of the mental let his view diverge from Ryle's behaviourism and in the end ultimately take a rather different course towards evolutionary thinking in general and 'teleo-functionalism' particularly.

2.1.3. Quine's Influences

Highly connected with the last point are the very influential works by Quine (1956) at that time, his idea of behaviourism and most importantly his work on radical interpretation. Quine's exploration of the concept of radical translation led him to the insight that strict reduction or translation of the idioms of meaning in general, and so the idioms of

intentionality into the language of the natural sciences is not possible. Starting from this point and conferring the insight to the discussion about a possible identity between mental and physical states, Dennett's refusal to accept any identity theory can be better understood, though Dennett's conclusions from the acknowledged fact of the problems of radical translation differed from Quine. Quine judged himself Brentano's irreducibility thesis as the mark of the mental to be highly related to his concept of indeterminacy of translation and as a reaction endorsed behaviourism, declaring intentional idiom to be baseless and empty (Quine 1960: 221). Dennett, although aware of the problems of radical translation, declared the sceptical part of Quine's position to be his own but refused to follow Quine in his behaviourism, noticing its unfulfilling prophecies. It has therefore to be kept in mind that though Dennett used to call himself a behaviourist on grounds of arguments that were put forward by Ryle and Quine, his acknowledging the arguments and problems is best interpreted as a starting point from which he developed an alternative middle ground between the proposed reductive and identity theories concerning the mental and behaviourism. Dennett's refusal to follow the behaviourists in their conclusions to the emptiness of intentional idiom or the proposal of mental states can be first understood as an answer to the general well-known problems of behaviourism, and second especially to the problem of normativity, deserving a further and separate discussion in the next paragraph.

2.1.4. Sellars and the Normative Dimension

Parts of Dennett's position with regard to intentionality have to be understood in connection to Sellars' theory of intentionality, aligning the use of intentional idiom with the concepts of normativity and prescription in connection with functional considerations (Sellars 1956). Dennett as well as Sellars recognize that intentional psychological idiom is not purely descriptive in its nature, but bears with it a normative dimension: beliefs can be *reasons* in normative space. Dennett holds it to be a critical characteristic of belief to be able to function as a reason in psychological understanding, and therefore any theory of intentional idiom to be in need of explaining this normative dimension. Behaviourism, as it was understood by Quine and Ryle, cannot do the job properly.

2.1.5. Predictability and Opposition to Cognitive Psychology

Dennett can be said to put special stress on the predictability of intentional attribution and looking at this characteristic as primary when compared to explicativeness. This point is best understood by considering Dennett's general idea of a philosophy of mind and its tasks.

As already mentioned, philosophy of mind is supposed to answer the most general questions about the mind and to provide a metaphysical foundation for the study of the mind in the natural sciences. Dennett's concern is hereby one of conceptual issues in the increasingly "mentalistic" and "cognitive" psychology of that time, e.g. the rise of cognitive science and the role of mental representation in cognitive psychology.

Furthermore, Dennett explicitly wants philosophy of mind to be a foundation for the reconstruction of ethical and moral theory, a corner stone of the concepts of personhood, responsibility, agency and morality, as possibly threatened by considerations under the heading of determinism in connection with materialism in terms of identity theories. Dennett acknowledges, with Fodor, the predictive strength of mentalistic idiom in everyday life, whose practise has been termed folk psychology. Dennett therefore follows Quine in at least one part of his "double standard" (Quine 1960), acknowledging the practical and pragmatic usability of intentional idiom, though contrary to Quine, not rebutting intentionalistic idiom in the end to be non-existent, but re-posing the question of its nature.

2.1.6. The Introduction of Rationality

Having observed the normative status of intentionalistic idiom and its pragmatic usability, as well as the need of interpretation of any usage of intentionalistic idiom, the question to the principles of interpretation and their presuppositions and implications as needed for radical interpretation must be posed. Influenced by behaviouristic thinking and sceptical of cognitive explanations as put forward by Fodor, Dennett answered the need by what he calls the Normative Principle comprising the Assumption of Rationality (Dennett 1971), i.e. the attribution of propositional attributes a creature ought to have. As a comparison, Dennett judges Davidson also to follow the normative principle in guise of the Principle of Charity (Davidson 1974). Rationality therefore in Dennett's account plays a two-fold role. First, it is used as an explanation to the origin of the normative part of intentional idiom, and second it seems to enable Dennett to leave the possibility of constructing a theory that can in the end vindicate personhood and responsibility, since rationality can be looked at as crucial to any ascription of responsibility and agent-hood. The exact nature of rationality in Dennett's work, however, is a highly complex and debated one.

2.1.7. Connections and First Conclusions

Taking these points together the stage for Dennett's intentional system theory is set. Following Quine, Dennett regards the attribution of intentional states as primarily pragmatic and interpretative. Acknowledging the irreducibility of intentional states and endorsing

procedural semantics, i.e. an analysis of the semantic properties of inner states in behavioural terms, Dennett puts forward a theory in opposition to Fodor's idea as merely repeating the question of radical translation, explicating and explaining intentionality in what I find to be a rather firm and cautious way. His aim is hereby naturalistic, fitting the mind in terms of its intentionality into the physical world. It is from here that the discussion of Dennett's work has to find a beginning.

2.2. Intentional Systems

Intentionality as used by Dennett has to be understood first as a theory explaining what it means to use sentences containing the idioms belief and desire. The kind of intentionality in question is the one often alluded to by the terms of "of-ness" or "aboutness", i.e. intentionality as it was introduced by Brentano (1874). Notably the starting point for Dennett's theory is intentionality of intentional idiom as exhibiting opacity in different contexts or "if substitution of co-designative terms do[es] not preserve truth or if the "objects" of the idioms are not capturable in the usual way by quantifiers "(Dennett 1971: 3). Dennett defines a system to which intentionality can be ascribed in the terms of belief and desire as an intentional system in so far as that its behaviour can be explained and predicted by relying on ascription of intentional states.

It has to be noticed therefore that on Dennett's account it is only possible to call something an intentional system, i.e. a system to which the possession of beliefs and desires can be ascribed in relation to an explanatory and predictive strategy of someone else. Besides the possible success of this strategy, there is no way to tell whether the subject of inquiry has "real beliefs¹" or not. There is no other way of having belief besides being attributed belief from the "intentional stance", where stance is a Dennettian term for predictive and explanatory strategy, a "heuristic overlay" (Dennett 1969).

The justification for the use of intentional idiom in prediction and explanation is purely pragmatic. As compared to the two other possible stances from which prediction and explanation of behaviour of a system may be conducted, namely the design and the physical stance, the intentional stance has both advantages and disadvantages. The advantage is its obvious high reliability and simplicity. Complex systems, may it be human beings or chess-playing computers, are so highly complex and the processes inside these not tractable with current technology, and for sure not tractable in everyday life, that the design or physical

¹ Until stated otherwise I will speak only of beliefs or desires to represent *all* propositional attitudes like desire, wanting or wishing, etc...

stance become pragmatically useless in prediction of their behaviour². Its disadvantage is the inability to explain error and malfunction of the system due to the assumption of optimal functionality³.

It is appropriate to use the intentional stance whenever it seems convenient, the physical or design stance is inaccessible and the assumption of optimal design is warranted. Its application does not depend in any way on the internal nature or kind of the object in question. The intentional stance assumes the system it observes to be first free of error and malfunction, and second rational. Rationality here means nothing more than “optimal design relative to a goal or optimally weighted hierarchy of goals” (Dennett 1971: 5). Prediction is relative to what the system knows, stipulated as what it “ought” to know and its desires are stipulated to be desires it “ought” to have.

The introduction of the normative dimension into intentionality is meant to do justice to the fact that beliefs and desires can be *reasons* for actions, and reasons are entities in normative space. The only anchorage for the ascription of desire and belief is the assumption of optimal design, the obvious need for survival and procreation and regularities in behaviour⁴.

To sum up, one predicts behaviour from the intentional stance by ascribing to the system the possession of certain beliefs, supposing it to possess certain desires and then by assumption of rationality, i.e. optimality, to predict what it will do by assessing what it “ought” to do. The decision to adopt the strategy is purely pragmatic, strategic and practical, not intrinsically right or wrong.

Dennett observes the intentional strategy to be deeply rooted in human practice. Having to interpret another system’s behaviour, we first assume its rationality and then adjust our estimate downwards when performance of the system reveals flaws. The more flaws one has to take into account, the more cumbersome the intentional stance predictions become and the

² Dennett’s special interest is hereby also to reduce the plausibility of reduction of mental states to neural correlates. As long as one is not able to observe the state of a complex neuronal system in sufficient detail, i.e. to track the states of a huge set of single neurons over time due to limitations in technology, according to Dennett arguments based on the assumption of such a possibility are better to be postponed in favour of arguments that do not make this assumption.

³ Of course it is also possible to judge from the intentional stance that somebody behaves queer, that his thinking is misguided or his conclusions wrong. The point of the hypothesis here is different. What *is* impossible to differentiate on the intentional stance is whether the system malfunctions and therefore makes a mistake which it would not make otherwise *or* whether the system is designed in such a way as that it just *must* malfunction.

⁴ If one follows this argumentation it might seem that all agents are to be ascribed the same desires and beliefs, since the only anchorage for ascription seems to be the same for all agents. To avoid this highly counterintuitive implication it is possible to observe that the assumption of optimal designed has to be relativized to the environment the system lives in, evolutionary history, its function and make-up (e.g. Footnote 7, Chapter 2.3.3). Thereby the assumption of optimal design for each agent becomes different and hence the counterintuitive implication is avoided. Since the status of the argumentation given here is controversial it will be led in detail in Chapters 5 and 6 of this essay.

more one is pressed to move to the design stance. This is a natural and helpful movement, since in the end only the design stance is able to explain the intelligence of the system in terms of its design. Predictions from the intentional stance from this reason always take a loan of rationality without being able to ‘pay it back, i.e. explain it by itself.

Since the success of intentional predictions is not based on any design consideration, but has a logical and normative basis, it cannot be used to confirm or disconfirm any particular idea of the system’s design. Dennett observes both folk psychological explanation as well as Skinnerian behaviourism to fail to provide a basis for scientific psychology and therefore proposes to look for regularities in human behaviour not in raw stimulus-response data, but in whatever mechanistic regularities there exist in the functioning of systems whose design approaches the optimal. Describing a problem in intentional idiom, one endows peripheral and internal events with content and then looks for mechanisms that justify the ascription of content.

Dennett’s notion of an intentional system is meant to be most basic and non-metaphysical, leaving out any consideration of consciousness, constitution, compositionality or morality. Whatever a person might be else, it is an intentional system in the first place. Further *possible* possession of other mental features, e.g. language or consciousness *presupposes* the system already to be intentional. If ‘mental treasure’ is to be purchased, it has to be purchased by the ‘intentional coin’.

2.3. Arising Problems of the Early Approach

As is clearly visible, in his early accounts Dennett has an *instrumentalistic* account of intentional states (Dennett 1971, 1981*b*), and this account has been subject to criticism in connection with different issues, most importantly Dennett’s aim to provide a proto-scientific basis for scientific psychology as well as moral theory.

2.3.1. General Scepticism Regarding Instrumentalism

The first and most obvious worry concerning Dennett’s theory can be found in a general scepticism concerning the aptitude of instrumentalist thinking. As Churchland notes, instrumentalism can be used to vindicate whatever theory one wishes to vindicate. A mere instrumental construal therefore cannot serve as a legitimising measure to hinder his own program, the “outright elimination” of the use of intentional system, i.e. folk psychology (Churchland 1981). This accusation is a general one, i.e. that intentionality instrumentally construed is worthless as any kind of basis for further philosophical discussion, reducing

intentionality to a mere chimera, or in Haugeland's terms "just a sham and a word play" (Haugeland 1982: 616).

Dennett does counter this attack. He holds that instrumentalist theory regarding propositional attitudes can do at least what he wanted it do, namely describe intentional idiom as a system for prediction. Since to him the existence of intentional idiom is vindicated already by pragmatic and practical considerations, i.e. its successful application, this kind of general criticism is too weak. As a *mere predictive strategy* instrumentalism regarding the intentional attitudes is immune to criticism.

Having ensured this point questions arise whether Dennett's instrumentalism does not only evade criticism with regard to its successful application to the intentional stance, but whether it can also be put to work in construing intentional attributions as apt to do the work Dennett wants them to do.

2.3.2. Criticism Regarding Explicativeness in a Proto-Scientific Theory

Dennett does not simply regard folk psychology as a candidate for transformation into an adequate scientific psychology, but rather scientific psychology to be two-fold- first, as foundational design/physical psychology⁵ and second as instrumental belief-desire psychology which is based upon the first. Dennett insists that on the one hand intentional ascriptions are only pragmatically justified and that on the other scientific psychology is to be found at the design or physical level (Dennett 1971, 1978f). Dennett's early position can hence be summarised as stating that true scientific psychologies are design/physical psychologies and that intentional folk-theoretic psychology is merely an instrumental myth, only useful for the prediction of behaviour. Reading Dennett's early theory we therefore get to know what it is to attribute intentional states to others, but we are not told *why* applying the strategy is successful. To put the matter in different terms, we are told what it is to believe that p, and what it is two persons believing that p have in common, i.e. being subscribed intentional states from the intentional states, but we are not told what it is that *enables* us to ascribe intentional states, and what is *responsible* for the success of the intentional stance besides its being applicable. Is there really nothing to the agent himself besides being the

⁵ It might seem that the concept of a physical/design psychology is incoherent, since pure physical or design descriptions do not capture the mental element of psychology at all. Dennett's account of the relation between the intentional and the design/physical stance explicated as the relation between abstracta and illata is supposed to be an answer to exactly this worry. Dennett himself holds a purely physical/syntactic psychology to be impossible. His idea in dividing up psychology in this way is to conceptualize the relation between the stances in such a way so as to avoid a pure physical psychology à la Stich (Stich 1983) and at the same time allow for some kind of connection between the stances. I bid the reader to wait for further explication of Dennett's theory on this point in Chapter 3 and 4.

subject of intentional interpretation that legitimises intentional attribution? What is being asked for is an explanation of why it is that we can apply the intentional stance under the assumption of rationality and optimal design. In the following I would to examine criticism on Dennett's position as it was posed by Lynne Rudder Baker (Baker 1989), Yu and Fuller (Yu and Fuller 1986) and others in order to elucidate Dennett's position regarding these issues.

Yu and Fuller recognize the main problem of Dennett's position to be the "mysterious" efficacy of the theory in question. As a matter of fact, according to Dennett to be an intentional system is nothing more than being interpreted from the intentional stance, but one may also say that it means to be interpretable from the intentional stance. As Dennett notes already in *Intentional Systems* (Dennett 1971), one does not have to assume the system one is interpreting to *know* the truth of logic, but to be able to *follow* the truths in order to explain the possibility to ascribe not only single beliefs to the system, but also automatically the beliefs that can be deduced from these. A *skill* is what is being asked for, but the possession of a skill is hard to be conceptualized otherwise than as a feature that the system in question possesses, and surely not primarily as an ascribed predicate. It is also important to notice that Dennett accepts assertions about the inner nature of the system only in so far as they follow from the assumption of rationality. Moreover, as Stich observes, Dennett accepts Ned Block's (Block 1981) argument that the internal nature of a system is relevant to the question of whether intentional ascription is warranted (Stich 1983: 244)⁶. These remarks contradict Dennett's comments on the purportedly non-essentialist nature of his approach.

If these inner features are thought of as beliefs there is no a priori problem for Dennett, since he is a revisionist with regard to propositional attitudes (Dennett 1981a). Therefore the question arises what can be said about the inner workings of a system interpreted from the intentional stance, and what the relation of the intentional stance to the physical and design explanations is, i.e. what the relation between being a believer from the intentional stance and being a believer on the design/physical stance is supposed to be. Dennett's early account seems to offer no answer to this question and puts pressure on Dennett that he tries to mitigate when developing his later theory.

To sum up, with respect to Dennett's early theory the problem arises that though the instrumentalist account is able to give a theory of folk psychology and also sees scientific psychology to be in the end grounded in the physical and the design stance, a possible connection between the two is not spelled out and no methodology is presented to figure the

⁶ Dennett also notes himself in 'Beyond Belief' (Dennett 1982a) that the assumption of intentional stance interpretability places limitations on inner mechanisms of the agent to be interpreted.

relation out. This failure can be looked on as what Fodor calls evading the “hard questions” while taking all the credit (Fodor 1985).

2.3.3. Criticism Regarding the Moral Project

The problem of the inner workings of an intentional agent recurs in the light of Dennett’s aim to provide a stable basis for moral responsibility. It is an obvious part of our concept of responsibility and agent-hood that the beliefs we possess play a role in our behaviour, i.e. *cause* our behaviour. This usage of the concept of causation needs further explication which will be given later (Chapter 6.2-6.2.3), but it should suffice for the time being that the kind of causation asked for is one that plays a role in sentences like “He did X, because he desired to...” or “He did Y, because he believed that p⁷”.

Beliefs, though, on Dennett’s instrumentalistic account can be criticized not to be able to be causally efficacious themselves, since they are thought to be merely *ascribed* entities from pragmatic and practical considerations, *not* internal states of a system.

Dennett is aware and convinced of the fact that intentional concepts are necessary in any account of morality, since being characterizable as a *believer* is necessary for being rational, hence being a moral agent at all (Dennett, 1971, 1983). Since these notions require causal efficacy Dennett requires some kind of explanation of how it can be that beliefs can play the causal role that they have to play in order to form a basis for moral agency. Dennett’s early theory leaves these issues open. Though Yu and Fuller argue that therefore Dennett needs non-instrumentally construed intentional states (Dennett 1978*d*), I propose to present a different argumentation in Chapter 3 and 4.

2.3.4. The Accusation of Irrealism

As a further point of criticism, there seems to be a danger lurking when it comes to interpretation and subjectivity of ascription of belief. One would rather want the possession of belief to be an objective fact than a subjective fact, since there seems to be more to having a belief than to be ascribed to have one. Beliefs are usually thought to be not just ‘in the eye of the beholder’ and the possibility that on ascription by one person the subject has a belief, and on ascription of another does not, seems counterintuitive.⁸ This point is important, again,

⁷ For an action to occur the agent has to *both* desire something and to *believe* that by performing an action he will get what he desires. In everyday speech though often only the belief or the desire is cited and the complementary desire or belief has to be inferred from the context. My example sentences are taken from everyday speech in order to stress the basic and common sense nature of the causality involved. Of course it is possible and formally necessary to complete the example sentences with the necessary belief or desire in question.

⁸ At this point it is also necessary to draw attention to what may appear to be a major drawback in Dennett’s theory (Hornsby 1997). Dennett’s intentional theory construes belief primarily as an ascribed predicate with

when it comes to belief as a foundation for moral philosophy and science. As Baker (Baker 1989) notes, Dennett construes belief instrumentally, but at the same time wants moral agents to “endorse” beliefs. The endorsement of intentionally ascribed belief can be criticised as at least highly counterintuitive, if not incoherent. Dennett’s instrumental position therefore seems to be rather unstable. The same criticism can be brought forward in connection with belief as a foundation for scientific psychology, since science is commonly thought to be concerned with objective, 3rd person facts about the world, not purely subjective interpretation as one fashions the world to be. To many critics Dennett’s intentional stance is not ‘3rd person’-enough, i.e. its intersubjectivity does not warrant objectivity.

2.3.5. Conclusions

It is as an answer to this kind of criticism that Dennett develops a theory of patterns, discernible in the world and interpretable in such a fashion that the ascription is supposed to gain objective status. The concept of patterns will be introduced and evaluated in Chapter 3. Dennett’s further solution is to start to spell out the relationship between folk psychological attitudes and what he gets to call sub-personal cognitive psychology. Hereby his task is twofold.

First, he has to give a *positive* account of the relation. Dennett accomplished the *positive* task by introducing the illata/abstracta distinction which I believe to evade Yu and Fuller’s criticism (Chapter 2.3.2). This topic will be discussed in Chapter 4. Dennett’s modified position hereby rebuts ontological commitment, becomes qualified, fills out explanatory gaps in the earlier account, and introduces instrumentalism of the intentional as what I propose to look at not as an *ontologically*, but *methodologically* necessary step.

predictive power from the third-person point of view, and *not* from the first-person point of view. On the one hand therefore Dennett’s theory in its foundations does not explain what it is for *me* to have a belief, i.e. what it means when *I* say “*I* believe that p”. To Dennett the status of conscious self-ascription of belief is merely derivative from third-person ascription of belief and therefore comes late in the formulation of his theory. Furthermore, to Dennett the categories of immunity from error and immediacy of first-person beliefs are *not* basic categories and are *not* thought to be the starting point for any kind of philosophy. His position on first-person beliefs is therefore highly influenced by his anti-Cartesian thinking. If data from a first-personal perspective is to be of any philosophical or scientific use, it must be gained by a revisionistic process termed heterophenomenology (Dennett 1991: Ch. 2-4, 1989, 2003). On the other hand it is not possible to state that Dennett neglects or leaves the topic out, since Dennett *does* offer an account of what it means for *me* to believe from a first person point of view. It is just that to Dennett any kind of self-ascription of belief has to be reconstructed from a complicated web encompassing a theory of third-person belief, a theory of consciousness as a virtual machine (and a theory of the self as a ‘centre of narrative gravity’ (Dennett 1990, 1992, 1999)). Due to the lack of space the special status of Dennett’s theory of first-person beliefs cannot be discussed in this essay. Short allusions to the topic can be found in Chapter 8 of this essay.

Second, Dennett also has to give a *negative* account so as to discourage belief in alternative conceptualizations, e.g. identity theories and functionalism. The *negative* task will be discussed in Chapter 6 in order to get clear in how far Dennett's position differs from other positions in the field.

3. The Later Dennett: Struggle with Objectivity

3.1. Objective Interpretation and the Pattern Theory

Among philosophers it is possible, roughly, to draw a division between those that construe ontological entities *instrumentally* and *subjectively* as objects of interpretation and those that construe ontological entities as *real*, as out there in the world independent of anybody postulating or interpreting them. Dennett, when it comes to the reality of "patterns" can neither be counted to one group nor the other simpliciter, but has to be considered to steer a *middle course* between these two. As belonging to neither of these groups, he is object of criticism to both kinds regarding his view either to collapse into interpretationalism or realism.

The first problem for Dennett is to evade radical interpretationalism (Rorty 1979) on behalf of saving intentional states like beliefs and desires to not only to exist 'in the eye of the beholder' (Dennett 1981*b*), foreclosing the possibility that one and the same object in question can be a believer from the point of view of one observer, but not from the other. Dennett proposes the following answer.

Though the attribution of intentional states is made by an interpreter from his own point of view, one ascription of a set of beliefs may have greater predictive power than another and thereby be more successful. Dennett argues that though interpretation may be subjective, success conditions are not. They are objective. In order to discuss Dennett's proposal I would like to split the task into two: first explain the case when one attributional strategy is superior, and second the case in which both are equal.

3.1.1. Superiority of One Attribution

If one of the sets of ascribed beliefs explains behaviour better than another, it is clear that the better set shall be used for pragmatic reasons. Dennett observes therefore that though ascription may be subjective, the success conditions of the ascriptions, i.e. how well the system's further behaviour can be predicted and generalized, are objective. The ascription

may even be relative to a culture, an interpreter's point of view or even personal likes or dislikes, constituting the subjective part of the stance, but how well the ascription works is not due to the subjective perspective of the interpreter, but to the objective success of interpretation. This point is important, since it is meant to explain away some of the awkward consequences of Dennett's 'pure' instrumentalism.

In discussing this observation I find it necessary to analyze the closely related topic of the *range of applicability* of the intentional stance first in order to give the discussion of Dennett's view of objective success conditions an appropriate form. The reasons for this need will appear in a moment.

Dennett stresses that the intentional stance can be applied to any kind of object in the world, may it be a thermostat, a lamp, animals or human beings. This position stroke critics as highly counterintuitive. They argued that the ascription of intentionality to a human artefact like a thermostat should be considered merely as as-if intentionality, whereas human intentionality obviously seems to be real intentionality after all. Dennett's position denies such an interpretation and Dennett offers instrumental, pragmatic and practical criteria instead that limit the applicability of the intentional stance. Though it may be that one could interpret any kind of object intentionally, the application of intentional attribution is warranted only if it is successful and pragmatically relevant, i.e. if it confers to the interpreter predictive power over and above what he can gain from the physical and design stance. Since we can explain the workings of e.g. a simple human artefact like a thermostat both in physical or design terms as well as or even better as in intentional terms, the application of the intentional stance is not intrinsically wrong, but in such a case pragmatically senseless. The application of intentional attribution becomes pragmatically legitimized when the objects in question show highly complex behaviour, such that tracking its design or physical states does not payout or becomes even impossible, as most remarkably in the case of human beings. The intentional stance thereby gains strength when the predictive aim is rather vague and broad, sufficiently reached by the exclusion and reduction of competing circumstances.

To offer an example, if I want to know which move you are going to do next in a game of chess, it is impossible to figure out this fact by assuming the physical or design stance, since neither is your design known to me, nor can I tract the processes in your physical constitution in any kind of pragmatically satisfying manner. The application of the intentional stance on the other hand gives me a good chance to predict at least roughly which moves you will favour. Thereby it is not important how exactly you hand moves or how the molecular make-

up of the chess board will be after you have moved your pawn, but what your move on an abstract level of chess moves will be.

As Robert Nozick (Nozick 1981) observed though, this explication *alone* does not vindicate Dennett's position after all. He argues that if human behaviour in principle were completely predictable from the physical stance without interpreting humans as intentional agents, one would be tempted to describe intentionality in purely relational terms. All there would be to intentional attribution would be relative to the interpreter. As it stands, this interpretation of Dennett's position points towards the more general conclusion that how strong Dennett ever could make his interpretationalism, as long as there were no objective matter to the fact in the world due to which intentional attribution were warranted, the problem of radical subjectivism could not be avoided.

At this point the strong connection between the concept of objective success criteria and the range of applicability of intentional ascription becomes clear. Nozick shows that the argument from the purported inapplicability of non-intentional stances *in some cases* to the necessity of the intentional stance fails. No objectivity that is not interpreter-relative and non-relative can be gained from purely pragmatic considerations. In other words, if intentional attribution can be substituted by observer-relative observation from the physical stance, the emergence of objective criteria is precluded from the very beginning.

Dennett's argumentation can therefore be malignly interpreted as not solving the problem, but moving it to the question of what warrants for any kind of objective criteria for success. This problem does exist for every stance, but since the intentional stance is the one from which semantic interpretation has to start, it is the intentional stance which faces the problem primarily. If one finds a warrant for objective criteria for the intentional stance, the warrant must and is being inherited by the lower stances⁹. What then, is the standard with regard to which one can judge one kind of interpretation more successful than another?

The answer to this question is highly complex, involving Dennett's treatment of the concept of rationality (Chapter 5). What at this moment shall suffice to notice is what the criteria *cannot* possibly be and to give a hint what they should be. Thereby has to be kept in mind that if a vicious circle of relativity is to be avoided it obviously cannot be the case that success conditions of belief attribution can depend on any kind of criteria couched in intentional criteria.

Dennett therefore proposes the concept of "*real patterns*" to steer a middle ground between interpretationalism and realism concerning the propositional attitudes. In doing so he

⁹ This hypothesis presupposes an explication of the relationship between the stances which will be discussed in detail in Chapter 4, especially Chapter 4.3.2.

makes use of a property of intentionality which is normally seen as highly problematic, namely the relation between its intension and extension. It is acknowledged that if one cannot appreciate that an unlimited number of possibilities can fulfil the success condition of an intentional predicate, one does not do justice to the nature of intentionality. As an example Dennett offers buying stocks at the stock exchange. There are infinitively many possibilities to buy stocks, e.g. by raising one's arm in one way or another, placing an order in a computer system or shouting. The number of physical states that can be described as "buying stocks" is infinitively big. Dennett states that if one does not see this fact, one has missed a "real pattern".

As follows from arguments for radical interpretation, the patterns to be discerned are never clear in a way that only one interpretation is possible, i.e. there is nothing intrinsic in one pattern rather than another that makes it the correct one. Still, it is a matter of objective fact when and why there is no fact to the matter. The objective presence of one pattern does not exclude the objective presence of another, and besides the success condition of adopting one or the other there is no matter to the fact which is the "correct" or "real" or "right" one.

Since the discernment and usage of intentional patterns is a *necessary presupposition* to be able to talk of any kind of intelligence, understanding etc. at all, Dennett rebuts Nozick's criticism. Nozick's approach presupposes that human behaviour *can* in principle be explained from the physical stance. Dennett in opposition believes this to be impossible. What Nozick therefore does in Dennettian terms is to miss a 'real pattern' and on these grounds his criticism can be uncovered to be inappropriate.

What there actually is in the world are an infinitude of patterns and observers of these patterns in tuned to these, selecting the one which can be discerned from their point of view and which ideally has the higher predictive power as compared to the adoption of another pattern.

3.1.2. Are Two Equally Good Interpretations Possible?

This observation brings one to the consideration of the second of two possibilities mentioned above, namely the concept of a case in which one interpretation, i.e. set of ascribed beliefs differs from another set which nevertheless exhibits the same predictive power. Looking at Dennett's position it is obvious that such a situation in principle may arise and that there is nothing intrinsic in one pattern or another which can make one interpretation the better one. When it comes to human beings, though, the possibility of radically different interpretations equally warranted by the intentional stance is pragmatically negligible. The kind of thinking warranting such a position is the following.

Interpreting a system intentionally we first ascribe to the system rationality, then beliefs and desires. Which desires and beliefs we attribute to the system is dependent on several factors. One is rationality, others e.g. are the desire to procreate and to survive. Furthermore, in the ascription of belief we attribute the beliefs which the system, given its constitution is most likely to have relative to the environment it lives in. A thermostat, therefore, can be attributed very different beliefs equally well, partly also depending on its environment. It can be said to believe the temperature of the room to be X degrees, or of the water which flows through it in a heating system, or of the human body whose temperature it is used to measure. In such a case there is *no matter to the fact* which belief the thermostat actually holds, and recalling the implications of radical translation this shall not strike one as surprising.

In the case of human systems this possibility of radically different but equally apt ascriptions of belief is *practically* usually reduced to one. Human beings as highly complex systems, occupying a complex environment and diachronically a specific evolutionary niche are highly unlikely to prove good candidates for multiple and radically different interpretations. Their involvement in the world and inner consistency of beliefs make radically different, but equally apt interpretations quite improbable. It is thereby important to notice how Dennett, though construing ascription of belief in behaviourist terms not unlike Ryle, can evade the impossible project of providing a behaviouristic determination for each single belief in isolation.

3.2. The reality of the patterns

I have argued that Dennett's theory of patterns was meant to foreclose the collapse of beliefs to interpretationalism as it was charged to do by both Rorty and Nozick, and so one may question whether his theory has been a full advancement or now in turn is prone to collapsing into physicalism. In evaluating the relation between Dennett's theory and physicalism I would like to examine Lynne Rudder Baker's (Baker 1989) critical discussion of the metaphysical status of Dennett's theory.

Baker recognizes Dennett as a realist about the physical, a position which Dennett himself subscribes to. The question arising in this context therefore is how the ontological status of patterns can be thought of. If there is nothing more in the world than the physical, one may conclude that the patterns must also be physical, that is describable as physical facts. Baker hence concludes that the difference between describing patterns from the intentional stance and from the physical stance amount to nothing more than a difference in use of *vocabulary*, where the one can be *reduced* to the other. If on the other hand patterns are not

facts describable in physical terms, what else can they ontologically be? If they are purely instrumentally construed entities, they are not able to play a causal role and Dennett's project of a foundation for science and ethics breaks down. Baker therefore notes Dennett to be in a dilemma: either there is something that eludes the physical stances from the intentional stance, that is just fiction, and hence cannot play its role, or the intentional stance is dispensable, since facts from the intentional stance can be explained as well from the physical stance in different vocabulary.

It has to be noted that Baker's criticism is similar to Nozick's, only that Nozick believed the elimination of the intentional stance to result in purely observer- relative facts, whereas Baker believes that only objective physical facts would remain. The difference in argumentation is fundamentally due to the respective philosopher's overall metaphysical position.

I propose not to follow Baker in her conclusions and despite of many other possible replies concentrate on Baker's equivocation of instrumentally construed entities with fictions. I am convinced that realism can be reconciled with the kind of instrumentalism Dennett suggests by the following argumentation. Dennett postulates beliefs as patterns in the world to be discernible from the intentional stance. Discernment thereby requires an interpreter and hence introduces a subjective element. This element can, according to Dennett, be cashed out in terms of success conditions of the interpretation, being in itself an objective fact. But what, then, does make the success conditions of the interpretation objective? Though the argument from the *mere presence of patterns* was enough to rebut Nozick's criticism, it is not enough to rebut Baker's. If there were *just* the presence of a physical pattern in the world, Baker's criticism would get a grip. It does not, because there is also *presence of an interpreter*. This interpreter himself is on Dennett's account of course made of physical stuff and describable in physical terms, but this is not the whole of the story. I stipulate that Dennett's point is not the mere presence of physical patterns in the world, but their presence plus the *presence of an interpreter* possessing the skill of attending to these patterns, being *tuned to* them and *reacting to* them. There are infinitely many patterns in the world and among these patterns *being interpreted* from the intentional stance.

Without the interpreter the patterns are still there, but they are not discernible in an "active way", i.e. they do not *play a role* in the sense of being useful. In order to be usable, to play a role in the prediction and explanation of subjects or systems, these patterns do have to be interpreted, have to be *used*. The physical facts about the patterns are there all along, as well as the physical constitution of the interpreter, but it is not possible to understand what it

means to *use* a pattern without the presupposition of somebody using them, i.e. the *interpreter*.

3.3. Wrap-up

To sum up, pattern theory does *not* collapse into the physical stance. Pattern theory also does *not* collapse into fictionalism and the realm of useful falsehood, since the physical facts about the patterns to be discerned are all there in the world, ready at hand to anybody who is apt and wishes to interpret them. I take for granted that the ontological status of patterns in Dennett's theory therefore can be guaranteed satisfactorily.

In turn, pattern theory does not come for free and obvious questions have emerged. What it is that makes an interpreter able to interpret patterns? What it is about him that warrants his being tuned and attending to these patterns? What it is about the system to be interpreted in favour of which it displays these patterns in the world?

Dennett offers a solution to this problem in terms of his intentional system theory and subpersonal cognitive psychology. It is an answer *both* to Yu and Fuller's criticism of Dennett's purportedly mysterious causal efficacy of beliefs (Chapter 2.3.2) *and* to the questions arising from the introduction of pattern theory. Explaining why pattern theory works in the end amounts to explaining how *interpreting* and *being interpreted* works. It is here that Dennett's theory of abstracta and illata has to be introduced.

4. The Later Dennett: The Relation between the Stances

As has been diagnosed in the last chapter, Dennett must offer a theory both to complement his pattern theory as well as explain the causal efficacy of beliefs. The main essays in which Dennett offers this theory are 'True believers' (Dennett 1981*b*) and 'Three Kinds of Intentional Psychology' (Dennett 1981*a*) In order to tackle the problems I therefore propose to follow Dennett in the exposition of the topics, first presenting a *modified version* of the instrumentalism and second discussing his ideas of subpersonal cognitive psychology in order to gain a clear understanding of Dennett's conceptualization of the relation of the stances.

To remind ourselves, Dennett follows Gilbert Ryle (Ryle 1949) in holding that conceptual answers are not to be given by micro-reductive psychology, since causal explanations of

phenomena do not coincide with naming a cause in the conceptual analysis (Chapter 2.1.2). Hence, Dennett asks which theoretical bonds are to be expected between psychological claims about intentional states and claims in biological or physical vocabulary.

4.1. Intentional system theory

According to Dennett the starting point for any kind of scientific psychology must be the folk psychological use of itself as a craft, the everyday concept of belief. This concept comprises the idea that beliefs are information bearing states of people, arising from perception and leading to action. Folk psychology does not only offer great predictive power (Fodor 1987), but also a misguided, incongruous part, i.e. superstitions, beliefs in the supernatural, prejudices, etc. (Churchland 1981) that hinders prediction. Explanations of this incongruous part may be important, but cannot supply one with a proto-scientific theory as a basis for scientific psychology. Dennett suggests to eliminate all there is in folk psychology that hinders its predictive power to arrive at a system that yield the greatest generativity and predictive success.

The outcome of such a process Dennett judges to be open, establishing himself therefore as a revisionist about intentional attitudes. Giving the example of ‘fatigues’, Dennett claims that there is no real entity, no natural kind that is being picked out by this term.

Dennett also observes folk psychology to possess several characteristics, first reliability, second generativity and efficiency, third easy learnability and fourth usability without any knowledge about the underlying mechanisms. Moreover, intentional description also bears a normative part in being reason giving relations, as already mentioned in the early theory, thereby making an ineliminable allusion to the rationality of an agent. The kind of rationality involved here is but quite different from what it was meant to be before. Normativity comes in as ideal use of an environmental niche, having the beliefs and desires one ought to have, that is being an ideal epistemic and conative agent, relativized to a set of needs for survival and procreation and to the environments his ancestors lived and adapted to, its perceptual abilities and its biography. Rationality is hereby not best viewed as an abstract notion of logical closure or consistency, but as a feature arising out of the adoption of a kind of adaptationism¹⁰. Dennett proposes to view this evolutionary thinking all the way in and therefore folk psychology to be understood best as a kind of *logical behaviourism*, i.e. to say

¹⁰ A full discussion of Dennett’s stand on evolutionary theory and thereby his special adaptationistic position cannot find a place in this thesis due to considerations of length. Parts of Dennett’s stand on these issues can be found in Chapter 5, esp. 5.1. The interested reader will find an exhaustive discussion of Dennett’s adaptationistic thinking in Dennett 1995.

that someone believes that p is to say that he will behave in certain ways under certain conditions.

According to Dennett folk psychology is *abstract* in so far as that intentional states are not supposed to be intervening distinguishable states of an internal system in the first place, but ascribed predicates. Dennett aligns them with *abstracta*, i.e. calculation bound entities like e.g. centres of gravity, and credits Reichenbach with this conceptual tool (Reichenbach 1938). At the same time Dennett observes that in a theory construing beliefs as the intentional theory does, the notion of belief causality remains unsatisfyingly vague. Therefore, being torn into two directions, first by the abstract nature of intentional states, but also second by their causal efficacy, Dennett splits the theoretical task into two parts: pure *intentional system theory* as just portrayed and *subpersonal cognitive psychology*.

4.2. Subpersonal Cognitive Psychology and Its Relation to Intentional System Theory

The task of subpersonal cognitive psychology is to discover the constraints on design and implementation variation and to demonstrate how individual systems realize intentional systems. According to intentional and evolutionary theory the brain hereby is a semantic engine, according to physiology a syntactic engine. Dennett acknowledges the impossibility to gain semantics from syntax, but observes that the brain can *approximate* the impossible task. All that natural selection can have produced is systems that seem to discriminate meanings (semantics) by actually discriminating something that co-varies with meanings (syntax). Sub-personal psychology generates and tests models of such activity. The postulated innards of the mechanisms can thereby be beliefs, but do not have to be; the only similarity we are sure to find is the “intentionality of their labels” (Dennett 1981a: 63)¹¹, characterizing states as content-bearing. To assign content it is necessary to understand the working of the system under *normal conditions* in its environment, since leaving out the environment either leads to unfruitful methodological solipsism or black world glass box perspectivism (Dennett 1981a: 64). The result in such a case would be abstract and purely syntactic neurophysiology. Dennett therefore notes, most importantly, that in such a case there can be no semantic interpretation, hence no psychology without intentionality. Intentionality is not only a possible, but a *necessary* starting point for psychology.

¹¹ A discussion of the relation of the syntax and the semantics of a machine couched in Dennett’s pre-illata and abstracta vocabulary can be found in Dennett 1978e.

Intentional system theory deals just with performance specifications of believers without caring about implementation. It is abstract and species unspecific, according to Dennett a necessary presupposition in order to say anything of interest about general topics like representation, intelligence and intentionality. Some kind of underlying structural processing is necessary to generate abstracta and their manipulation, and the realization will surely resemble the complexities of the instrumentalistic interpretation strikingly, but due to evolution and difference in biography it seems illusory to expect the same mechanisms in any two intentional systems¹². General prospects of micro-reduction of psychology to physiology are therefore nought.

A possible prospect, though, is the reduction of folk psychology to intentional system theory. According to Dennett Brentano's irreducibility thesis is answered by a reduction if every mental predicate is intentional-system characterizable. To explicate the relation, other than reduction, between intentional system theory and subpersonal cognitive psychology Dennett invites us to accept the following conceptual distinction: distinguish *illata* from *abstracta* (Reichenbach 1936).

Illata are posits of a theory, whereas abstracta are construed bound entities (e.g. centres of gravity), not reducible and causally efficacious. Dennett aligns revised folk theoretical intentional states with abstracta. He is hereby not interested in folk psychological 'impure' beliefs which have a status between illata and abstracta, but beliefs as pure abstracta, being the subject matter of decision and game theory, which are 'similarly abstract, normative and couched in intentional language'. Subpersonal intentional states postulated by an as yet uncompleted subpersonal cognitive psychology on the other hand are genuine illata.

Illata as posits of a theory are to be gained in the following way, mirroring the standard methodology of cognitive psychology followed knowingly or unknowingly at that time. Starting with an intentional system, one constructs intentional models according to the top-down strategy by deconstruction into smaller systems. The important point as Yu and Fuller note, is that each subsystem is characterizable in intentional terms as content bearing states, albeit more "stupid" than the systems up in the hierarchy¹³. These postulated entities may, though of course do not have to resemble beliefs or desires except in being intentionally labelled. Arriving at such a low level, we may characterize intentional systems without strain

¹² An argument explaining this hypothesis regarding the relation between abstracta and their underlying mechanisms can be found in Chapter 6.1.

¹³ The hierarchical structure of gradually growing 'stupidity' in the decomposition of intentional systems forms the basis for Dennett's account of the origin and genesis of intentionality as discussed in Chapter 7.2

in mechanistic, physical terms¹⁴. To put the matter in a nutshell, genuine illata therefore will be subpersonal intentional states, the theory concerned with realization of intentional systems 'subpersonal cognitive psychology'. This two-tier system of intentional system theory/subpersonal cognitive psychology is analogue to the old two-tier system instrumentalist superstructure/design physical substructure.

Intentional systems are thereby realized by physical systems in the sense that first the abstracta of intentional system theory are defined in terms of genuine illata in subpersonal cognitive psychology, and second illata in subpersonal cognitive psychology just *are* states of physical systems, intentionally characterized. In this context Yu and Fuller put forward the thesis that Dennett seems to accept the following distinction. A theory is *instrumentalistic* if the terms of the theory are not claimed to refer and the sentences of the theory are not literally true, though the theory is claimed to have predictive utility. This is how Dennett viewed folk psychology in his early writings. A theory is an *abstracta theory* if the terms of the theory are claimed to refer to calculation bound entities and the sentences are claimed to be literally true. Terms and sentences are shorthand for terms and sentences in a full-blood realist theory. This is, according to Yu and Fuller, Dennett's later position. A theory is a *realist theory* if the terms are claimed to refer and the sentences are claimed to be true. This is Dennett's view on subpersonal cognitive psychology.

If Dennett's construal of the relation between illata and abstract is successful, Dennett has produced a realistic foundation for personhood and responsibility, also vindicating folk psychology. At the same time he can reject Fodor's thesis that ordinary folk theoretical posits will turn out to be genuine illata in an adequate scientific psychology. Furthermore, he can claim to have explained causality at the level of belief in relation to subpersonal states, provided non-reductive explanation of belief, modified his instrumentalism so as to evade gross irrationalism and enriched his account successfully.

By doing so he could hold that he has answered the questions posed in the end of the last chapter. By providing us with an explication of the relation between the intentional stance and the design stance in terms of illata and abstracta he has explained the workings of an intentional system and thereby also why these exhibit patterns which can be interpreted intentionally.

Naturally, in reality nothings turns out *that* nice and Dennett's construal has not been received favourably. First, Yu and Fuller have criticised Dennett for construing the relation

¹⁴ The term mechanistic is hereby not to be conceived of too narrowly, since subpersonal cognitive psychology is interested in the realization as well as *realization constraints* of intentional systems. Not only the actual physical structure of the system is in our interest, but also more general *specifications* on the physical build up.

between abstracta and illata in a fashion too loose to be of any use. I will discuss and rebut this accusation in Chapter 4.3.1. Second, Yu and Fuller identify a problem in Dennett's illata/abstracta distinction with respect to normativity. The evaluation of this problem will be done in Chapter 4.3.2. Moreover, its discussion will lead to the recognition of a major and general problem in Dennett's account.

4.3.Criticism

4.3.1. 'Too Great a Leap'

Dennett argues for folk theoretic intentional states being abstract by the following inference. Intentional states are ascribed to an intentional system as causally efficacious states, but they are not to be identified, i.e. reducible to internal content bearing states of the system. Therefore they are abstract, idealized entities and shall be aligned with theoretical, logically bound entities in Reichenbach's sense as abstracta.

Yu and Fuller (Yu and Fuller 1986) criticize that on such a construal neither personhood can be founded, nor psychological notions be vindicated any more than through a purely instrumentalistic theory. First, they note that it is unclear whether abstracta like centres of gravity are any more causally efficacious than instrumental states. On Dennett's new account intentional states are abstracta and real in the sense that their ascriptions can be true, but not real in the sense of causally efficacious internal states. But since abstracta are calculation bound entities, they must be *definable* in terms of illata, i.e. the posits of the theory. If so, they argue that Dennett's inference is an incredible leap, since expressions designating abstracta are introduced late when the theory is already explicitly formulated. Whether folk theoretic intentional states are abstracta depends on the nature of a *yet unknown* subpersonal psychology. Yu and Fuller argue that there is no reasonable argument favouring such a psychology and therefore Dennett's argumentation is rather stipulating than convincing.

I doubt that Yu and Fuller's critique is as strong as it seems to be and whether there is any *special* problem for Dennett here. On Dennett's account, beliefs as intentional states are *not reducible* to inner states simply, but the connection is more complicated. Dennett argues for this thesis by noting that as it comes to the relation of the stances, the only thing one is sure to find on the subpersonal level are core elements that have an *intentionalistic label*. No mirroring of the causal structure at the abstracta level is to be expected at the illata level. Yu and Fuller argue from this that Dennett's account is empty, but I cannot see this to be so otherwise than on the prior assumption, that the causal structure at the abstracta level will be

found again on the illata level. Furthermore, I do not see that Dennett's account presupposes this at any point or that this presupposition in general is warranted. Therefore, Dennett's "leap" from abstracta theory to illata theory is huge, but not bigger than on any other account in general. It has of course to be noted that Dennett's position as it comes to the vindication of intentional states is not as strong as to silence any doubt, since Dennett is himself a revisionist about the intentional states in the first place, not a vindicator.

Can on this ground be said that Dennett's vindication of the intentional is endangered? I don't think so. Assuming even that during further work in cognitive psychology no vindication of intentional states as discrete internal content bearing states could be made, such a result would not endanger the plausibility and strength of Dennett's position in any way. To remind ourselves, if we follow Dennett in believing that intentional states are posits of a rationalized and abstract predictive theory without any presupposed limitation on the exact realization of these states, the way these states will be realized is no matter of interest. The only important point is that subpersonal cognitive psychology will have to explain how it comes about that systems are interpretable as intentional systems, and nothing more. To ask for further explication of the connection between subpersonal and personal states on philosophical grounds is to ask too much. Beliefs as abstracta are to be defined in terms of yet unknown illata, but since no one-to-one correspondence between beliefs and internal states is to be expected, Fuller and Yu conceptualize the notion of definition in terms of direct reduction too narrowly. I propose to follow Dennett and conceptualize the definition of abstracta broadly in terms of illata through a process of vindicative interpretation and scientific discovery following the method of intentional labelling.

On these grounds Dennett's ontological commitment can be observed to be strongly shaped by his epistemological assessment. Possessing no alternative way of describing systems as intelligent and intentional in intentional terms, the methodological starting point for scientific discovery has to be intentional labelling of whatever one may find to have labelled and whatever turns out to be there at all.

4.3.2. The Problem of Rationalization

Yu and Fuller further observed important criticism of the status of illata and abstracta with regard to rationalization and optimality. In his early theory Dennett didn't treat beliefs as illata, because though physical and design theories are to be true of systems, correct intentional theories will *never* be true of systems due to rationalization. Furthermore, Dennett argues that it is highly improbable that three people sharing the same belief will have a 'similarly structured object in each head', and many folk psychological entities like fatigues

will not be found either. Yu and Fuller argue that these points taken together prevent beliefs not only to be *illata*, but also *abstracta*, since if subpersonal psychology is supposed to be literally true and beliefs calculation bound entities, then beliefs would have to be things literally being possessed by certain systems. On the other hand it is hereby important to notice that whatever rationality is in Dennett's opinion, it is always idealized and object to norms.

I believe Yu and Fuller's point to be correct in some regards, but false in others. As they note, Dennett's notion of *abstracta* is highly rationalized and on first view it seems impossible that rationalized and idealized entities can play a role as thing literally in the head of a system. If Dennett's view were to warrant such a conclusion, the criticism would stick. I would like to argue that this is not the case by the following short argument

Although Dennett construes *abstracta* as rationalized and idealized entities and *illata* as entities that *prima facie* inherit these characteristics, Dennett puts great stress on the fact that realized intentional systems are not semantic, but syntactic machines. All that these machines do is to *approximate* a task and to work *sufficiently* well, not optimally. To reconcile the contradictory views on intentional systems and their rationalizations it is necessary to see that though Dennett makes the assumption of rationality and optimality, he does so *epistemologically or methodologically* and not ontologically. One possible answer therefore is that the *method* to discover *abstracta* and *illata* makes rationalization and optimal design necessary in order to account for the normative dimension of reason and to enable a semantic description of systems in the first place, but does not commit to the assumption of optimal and semantic realization at the ontological level.

Unluckily such an argumentation in favour of Dennett may be too weak, since to many philosophers the gloss of a position as methodological is just an illegitimate argument to hide the *metaphysical* instability of the position. Dennett's own position on the relation between metaphysics and epistemology is rather vague, consisting mainly in the remark that epistemology and metaphysics do go 'hand in hand' (Dennett 1996). It can also be observed that Dennett's project of naturalizing the mental is not to bring our concepts of mental states in accord with our concept of the physical, but to convince the reader that the concept of the physical has to be changed in order to encompass the mental. Dennett's answer to the accusation of metaphysical instability would therefore be that this only shows one's metaphysical concept of the physical to be wrong¹⁵.

¹⁵ In Chapter 7.2 I put a part of my explication of Dennett's position with regard to rationality of design as compared to rationality of single actions also in terms making use of the epistemology/metaphysics dichotomy. The same arguments for the applicability of this dichotomy also apply there. If one feels convinced by my argumentation, my interpretation of Dennett can be regarded as showing one of the underlying motivations for Dennett's position. If one does not feel convinced, it does not matter much, since the arguments in Chapter 7.2

Another possible and related answer involves Dennett's concept of *rationality*. In short, it is to state that rationality is not to be construed as deductive closure or logical consistency, but as what is best from an evolutionary point of view. Rationality is to be construed differently than Yu and Fuller do and therefore to evade their criticism. The kind of rationality involved and necessary is an attenuated notion, tinted rather by evolutionary thinking than by logical standard. Unluckily this kind of argumentation cannot be taken seriously without a detailed discussion of Dennett's concept of rationality, since it has to be shown *how* the problematic notion of rationality can be construed differently so as to foreclose Yu and Fuller's criticism.

I therefore propose to discuss Dennett's concept of rationality next in order to flesh out what exactly Dennett has in mind speaking of rationality. Getting clear on this issue is necessary to get a clearer picture of Dennett's illata and abstracta concepts and show whether Dennett's theory can be vindicated. I will argue that it cannot do so for free, but on cost of necessity of further development of Dennett's theory. Thereby the discussion will also open the space of what I will hypothesize to be a new strand in Dennett's philosophy of mind, different in emphasis and character as compared to Dennett's earlier writings.

5. Rationality and the Evolutionary Turn

5.1. Introduction, Early Position and Criticism

I have been arguing that rationality does play a major role in Dennett's position. In order to explain any system from the intentional stance, we have to attribute rationality to it. Dennett's comments on what rationality is, i.e. what it is that we ascribe to systems every time we use the intentional stance, has been very scarce. It is therefore hard to get clear about the concept of rationality Dennett uses and how it developed. I would like to argue that Dennett's view on rationality has changed over time in explication and as an answer to criticism. Further I would like to hypothesize that understanding the development of Dennett's concept of rationality is crucial to understanding the reasons and the basis for Dennett's newest philosophy of tools, language and memes.

In his early "Intentional Systems" (1971) Dennett aligns rationality with optimality, claiming that the most rational action of an agent is his best action, and that choosing the most

do not rely on the correctness of *this* part of my interpretation. Moreover, one can find Dennett's comments on his conceptualization of epistemology and metaphysics in opposition to Jaegwon Kim (Dennett 1993).

rational move is equal to the system's design being optimal. As Dennett further notes, rationality to him here means nothing more than optimal design relative to a goal or optimally weighted hierarchy of goals and a set of constraints (Dennett 1971: 6).

When we set out to interpret agents, we start with assuming rationality and then learn to adjust our estimate downwards when the system performs badly. Our assumption of rationality is said to be deeply entrenched in our practices, such that we are more likely to assume perceptual errors or goal weightings before we assume the agent to be irrational. Furthermore, Dennett notes that since the assumption of a system being intentional is the system being rational, we have to attribute beliefs to the system *and* the beliefs that follow logically from these. I.e. even if we do not assume the system to know the truths of logic, we expect it to follow the truths of logic (Dennett 1971: 11). Dennett stipulates that if a system were perfectly rational, it would follow all truths of logic, but since we can expect all systems to be merely imperfect, not all rules will be found. Moreover, one expects to find imperfect rational creatures whose allegiance to modus ponens varies with the subject matter, and therefore excludes modus ponens as a rule and ascribe instead a set of non-logical inference rules.

Dennett notes that as one discovers more and more of such failings, the application of the intentional stance becomes *more and more* "cumbersome", leading *in the end* to the adoption of the design stance instead and dropping of the assumption of rationality. Dennett observes this to be a natural movement leading to explanations of the systems design and in turn of the design by natural selection.

I would like to draw attention to several points concerning Dennett's construal of rationality in his early phase. First, Dennett notes rationality to be a *gradual notion*, i.e. creatures can possess more or less rationality and there is no single limit, no line that has to be crossed for being a rational creature. The upper limit to rationality, though, is the possession of exclusively true beliefs, the optimal weighting of goals and the ability to deduce all and only all correct beliefs that follow from the ascribed "core beliefs". Although Dennett assumes that perfect rationality will never be found in any living creature, he still proposes that we start interpreting agents by assumption of a full-fledged rationality. If this is so, one can observe that how much rationality and what kind of it Dennett wants us to attribute to intentional systems remains unclear. On the one hand we are to ascribe pure rationality in logical terms, on the other he observes this assumption to be unwarranted.

Dennett's notion of rationality construed as pure rationality in logical terms therefore is unconvincing. First, Dennett's notion of rationality might be way too strong to be of any role

in explaining the nature of even what we paradigmatically expect to be the most rational beings, namely ourselves. If we assume rationality to consist in the ability to deduce all possible true beliefs from a set of already true beliefs according to the rules of logic, one may gain the impression that what Dennett means by rationality is deductive closure and logical consistency. If this is so, then such a notion of rationality is too demanding. It is not the kind of rationality that any kind of agent build out of mechanistic parts could ever have, unless it is a very simple system in a very simple environment. Furthermore, no human being does possess only consistent beliefs, but also inconsistent ones, without us wanting to call him irrational on this account.

Moreover, as Stich observes (Stich 1981), one runs into problems explaining human behaviour as rational in Dennettian terms that are described as rational usually in everyday life. Humans make decisions, act and form beliefs in a way that cannot be described as following the rules of logic. We often use short-cuts in our decision procedures before finding the optimal solution and neglect beliefs in favour of others without evaluation. Nevertheless such procedures are rational, or at least more rational than attaching oneself to the truths of logic.

Second, on the other hand Dennett's notion of rationality may be way too weak. Dennett claims that the best move under certain circumstances as a result of the computation in the system's design that is supposed to be optimal is also rational. As Dennett puts it, rationality is the design stance plus the assumption of optimal design. Furthermore, the only explanation for the existence of design in nature when it comes to living beings is natural selection.

The implications of such a view are striking and endangers Dennett's aim to keep the design and the intentional stance apart. If natural selection is the standard against which rationality is measured, the notion of rationality collapses into whatever natural selection brings about, telling us that we could never call any organism irrational. Furthermore, if rationality is supposed to be what evolution brings about as a best solution, we may expect imperfect logical thinking to exist in the creature as a result of being 'evolutionary' rational. This point can be argued for by different strands. First, one can argue that an *apparently* irrational procedure is 'the best' or 'optimal' procedure that a creature build out of mechanistic parts can achieve. Second, one can argue that an *apparently* irrational procedure may be close to the best or optimal procedure that an agent designed by natural selection could achieve.

5.2. Dennett's "Answer"

Dennett's initial notion of rationality is instable, collapsing either into an improbable standard of deductive closure and logical consistence, or into an uninformative notion of rationality as whatever evolution gives us. Dennett in his later phase does not mend this fact directly by giving us a clear definition of what he means by rationality, but rather what he does *not* mean by it. In 'Making Sense of Ourselves' (Dennett 1982*b*) he states, that

"[Rationality] is not deductive closure...Nor is rationality perfect logical consistency...[And] I am careful not to define rationality in terms of what evolution has given us...What then do I say rationality is? I don't say... (Dennett 1982*b*: 94-96)".

The closest Dennett comes to a definition is stating:

"I want to use rationality as a general purpose term of cognitive approval- which requires maintaining only conditional and revisable allegiances between rationality, so considered, and proposed methods of getting ahead, cognitively, in the world. I take this usage of the terms to be quite standard, and I take appeals to rationality by proponents of cognitive disciplines or practices to require this understanding of the notion (Dennett 1982*b*: 97). "

Dennett's claim here can best be described, as Elton notes, that it is possible to "work with the concept of rationality as cognitive approval and get all the results from his account that we need (Elton 2003: 128)". As long as we can reach some kind of agreement in some cases about what is rational or not, and this is possible since such a concept of rationality is quite standard as Dennett notes, and 'is' implies 'can', we do not need any a more detailed account.

Dennett's answer to Stich's challenge that humans are often governed by psychological principles that are not logically rational and which Dennett does not seem to accommodate can therefore be explicated in two ways. On the one hand, as already stated above, Dennett ascribes the predicate of rationality to *single actions* in certain circumstances, i.e. the move that is the best or optimal action given circumstances, beliefs, goals and information is the rational one. This explanation is a personal level explanation, making no allegiance to the system's design or history. Mistake or systematic failure is hereby explained against a background of success.

On the other hand, Dennett also tends to shift away attention from particular action to the *design* or history of the system. It is here that the arguments that an *apparently* irrational procedure is in reality rational nevertheless applies because it is the best that agents build up out of mechanistic parts can do, or the best result that a creature shaped by natural selection

can achieve. Dennett shifts the concept of cognitive approval away from particular actions onto the design of the agent and thereby into the subpersonal level¹⁶.

5.2.1. Rationality as Applied to Single Actions

If one follows Dennett's notion of rationality as applied to single actions, one can explain failure in perfect procedure and at the same time legitimize the ascription of rationality by stating that human beings can accommodate a large number of non-rational principles provided a *rational background*, i.e. a real, though noisy pattern of rationality is preserved nonetheless. In this way Stich's criticism can be rebutted. This kind of explanation draws on Dennett's pattern theory and reminds us of the fact that a large amount of rationality has to be assumed to be in play to even make sense out of a subject's utterances in the first place (Dennett 1991: 72-8). Talking to agents and expecting them to understand is to ascribe to them a great deal of rationality already. Since therefore slips in procedure or irrational behaviour of agents do happen against a background of behaviour that conforms to rational norms, there is no reason to give up the ascription of rationality, as Stich seems to hold to be unavoidable on Dennett's account. Elton judges this kind of response to be the better one, but I cannot follow his conclusion. Though in itself there is nothing wrong in viewing Dennett applying rationality to single actions, this argumentation cannot answer the questions emerging out of Yu and Fuller's criticism (Chapter 4.3.2). Conceptualizing rationality as rationality of single actions there is no possibility to apply the notion to illata and abstracta. Abstracta are calculation bound entities, and illata posits of a subpersonal psychology. How could rationality conceived as rationality of single actions be understood to play a role here or to be applied? I do not see any meaningful connection. Conceptualization of rationality as rationality of single actions *cannot* be the solution of an alternative conceptualization of rationality that Dennett needs.

5.2.2. Rationality as applied to a System's Design

I would like to argue that if one conceptualizes Dennett's notion of rationality as applied to the system's design one arrives at better solution. It is better first because obviously it is able to make direct statements about the rationality of illata, since illata directly refer to parts of a system's design. Second, one arrives at a notion of rationality much more oriented towards a deflationary account of rationality and so better suited to fit Dennett's overall naturalistic aim. Evolutionary thinking is the more 'down-to-earth approach' as compared to a

¹⁶ To foreclose misunderstanding it is important to notice that saying that certain procedures are the best procedures that could be installed is not the same as saying that certain procedures achieve the best result given the circumstances (Elton 2003: 129). The argument here says the former.

more abstract assumption of mistakes just making sense against some kind of background of rationality.

Stich's argument from 'rational error' is handled in the following way. A mechanistic agent is not able to fulfil the assumption of deductive closure, since it is limited in its capacity to deduce all possible beliefs. Mechanisms do not come for free, but cost both time and energy to construe the necessary machinery and to run it. It is therefore not rational to consider every implication, since time may be too costly in a given situation. If one leaps to a conclusion by figuring fewer implications, one takes the risk of missing something important, but calculating too long might be life threatening, e.g. when the task at hand is predator detection (Dennett, 1996*b*). Furthermore, it might be better to make more misidentifications or judgements than fewer, e.g. in the case of assessing an enemy's strength as bigger than one's own. Assessing one's own power too low may cost the system a single possibility to gain food or to procreate, but assessing it too high may cost one's life. Therefore we are willing to give cognitive approval to a system that leaps to conclusions from time to time or misjudges situations. Rationality is a trade-off notion between 'looking ahead' and 'action'.

The second argument that an logically irrational procedure may be the best or optimal procedure that an agent designed by natural selection can achieve can be explicated in the following way. Evolutionary thinking places a number of constraints on design processes. Before all it has to be observed that evolution does not deliver optimal, but just satisficing designs. Good or best means here 'good enough' to survive and procreate. Following Elton I would like to discuss two further constraints in this regard.

First, all new designs have to be built on already existing designs, so that evolution is a gradual process and not a process of leaps from one optimal design to another. A new design originating from the old one is limited to the prior calculation capacity of the old one and the design of the new one is therefore not arbitrary. Evolution does not design the best solution to any given problem, but the best solution to any given problem given the history of the system as implemented in its already existing design. Even if another design could be much more successful, if it cannot be created by gradual steps from the old one, it cannot come into existence. To illustrate the point one might draw an analogy to the developments of planes and the change from propeller machines to jets. Though a jet compared to a propeller machine is surely a more optimal solution¹⁷, it seems impossible to envision how the change could

¹⁷ Of course in some cases a propeller machine is better than a jet, e.g. when one has to land the plane in the woods. In *some* aspects a propeller machine is more optimal than a jet and vice versa. This being so nevertheless does not make Dennett's point weaker, since *any* kind of measurement of optimality is *relative* to some set of goals or other. Relativity itself is a different issue dealt with in Chapter 8.2.

have been brought about by natural selection. It is impossible, since the number of gradual steps from one design to the other, where each step is to be a better design than the one before, is hard to envision. Nonetheless in the case of design by natural selection such a process has to be imaginable. To draw the analogy further, one may want to say that a jet is more rational than a propeller engine, but one should not expect from this reason to find the development of jets from propeller machines a natural development by natural selection.

Second, evolution is not able to look ahead. It does not calculate what could be the best solution to a problem and implements it, but simply chooses between existing possibilities due to their differential success in procreation. One is not to suppose the best solution to emerge, but the best possible given the alternatives.

Due to these considerations purportedly logically irrational procedures are to be considered and described from an evolutionary point of view as rational after all. We should therefore give cognitive approval with respect to the system's design and evolutionary history, rather than to individual actions of an agent. We might give cognitive approval to a system's irrational behaviour if it is observed in circumstances it was not designed to behave by arguing that in its own environment it does behave rationally. Such a reading of Dennett is in accord with his sympathy for Minsky's multiple agent model of the mind (Minsky, 1985) and can also be observed to be in accord with other evolutionary theories of cognition. As Cosmides and Tooby (1992) note, human intelligence is best interpreted as a bag of instincts designed to the solution of particular problems rather than as a general problem solver. The view of the human mind as a general problem solver can be aligned with the assumption of deductive closure and logical inference patterns, whereas the bag of instincts is much better explained by step-by-step satisficing particular problem solving routines designed so solve particular problems. If one adopts the more evolutionary tinted approach to rationality, one can mitigate Dennett's earlier claim that by assuming pure rationality and discovering rules of logic, e.g. *modus ponens* to be limited to a subject matter, one would have to step down to the design stance, and hold that one can remain on the intentional stance even if rules of logic are not universal.

5.3. Wrap-up

To sum up, as I have tried to argue, Dennett's notion of rationality has undergone changes in terms of emphasis from his earlier position to his later one. The assumption of rationality is best understood as primarily a feature of a system's design and secondarily as a

property of a system's particular actions in particular circumstances. In this form it can both foreclose Stich's as well as Yu and Fuller's criticism.

For a system to be intentional is for it to be rational, so that this chapter could equally well be read exchanging the word rational by intentional. It is thereby important to notice that Dennett's explication of intentionality has taken an *evolutionary tint* and a turn away from the personal level to the subpersonal level of the design stance. I propose that it is from this philosophical position that Dennett's newest writings on the *design of intentional systems* can best be understood.

Dennett's adoption of meme theory, his theory of evolution of the human mind as a very particular kind of intentional system and finally his account of the human mind as a bag of tools as an explanation of what a system and especially a human system needs to be to enable it to be the intentional system it is can therefore be best looked at as possible extensions of his theory of intentionality. It is in this typically Dennettian, i.e. partly conceptual and partly empirical spirit that I suggest to advance in Chapter 7 and 8.

To remind the reader, before doing so it is necessary to finish the discussion of Dennett's take on the relation between the stances by giving an account of Dennett's *negative task* (Chapter 2.3.5) of discouraging belief in alternative conceptualizations of the relation between beliefs and physical states of a system in the next chapter. This discussion will serve two aims.

First, it makes clear in which ways Dennett's position differs from other philosophical positions in the field and thereby allows localization of Dennett in a philosophical landscape. It thereby helps to shed light on Dennett's *unique conceptualization of intentionality* and its role as compared to its conceptualization in other prominent approaches often *compared or confused* with Dennett's. This task will be tackled in Chapter 6.1. In Chapter 6.2 it is continued by comparing Fodor's and Dennett's position. As Dennett most often has been claimed to hold a functionalistic position, a comparison with a typical functionalist like Fodor will clarify the differences between the approaches exemplarily. Moreover, in the debate of the differences between the approaches observations important for understanding of the form of Dennett's newest writings will emerge.

Second, together with the results of Chapter 4 and 5 the explication serves as an explanation of Dennett's *motivation and reasons* to rebut belief in the possibility of micro-reduction of psychological states of any kind and to turn to macro-reduction instead. Dennett's turn to macro-reduction of intentional states will be the topic of Chapter 6.3. Since Dennett's rebuttal of micro-reduction and the subsequently mandatory turn to macro-

reduction are responsible for the content of most of Dennett's newest writings on intentionality, a discussion of the reasons and nature of this turn is indispensable. Since the background of Chapter 6 should have become clear, I propose to jump right into the discussion.

6. Micro- and Macro-Reduction

6.1. Dennett and His Place in the Philosophical Landscape

To locate Dennett's position in the philosophical landscape I propose to argue in the following way. Since Dennett holds a naturalistic position, his rebuttal of any kind of substance dualism follows directly. Moreover, Dennett does not hold any of the other naturalistic positions that were put forward in the last fifty or sixty years. He rebuts identity theory as type physicalism, i.e. the claim that for every mental predicate there is some other predicate expressible purely in the vocabulary of the physical sciences¹⁸, on grounds of holding necessary and favouring a *functionalist description* of systems over a physical one.

Dennett's way to state his position is not to give formal arguments, but to illustrate his points by examples to make the reader see how unfortunate the criticised position and its implications are¹⁹. The example Dennett favours in order to discredit type identity theory is to understand that as it comes to objects like e.g. clocks or can-openers, which are beyond doubt purely physical things, it is not even remotely possible to suppose that one could compose a predicate in the vocabulary of physics that singles out all and only clocks and can openers in the world. Dennett draws attention to the fact that what clocks have in common is a *purpose or function* regardless of physical constitution. In this regard Dennett's argumentation is functionalistic in the sense of being based on appreciation of some concepts as being *teleological*.

¹⁸ This gloss of type physicalism may seem astonishing to the reader which correctly may want to gloss it better as semantic physicalism. Dennett's equivocation of the two positions is best understood as arising out of the traditional way in which the problem of intentionality was conceived, i.e. as a problem of the relation of mental *language* to physical *language*. Today the positions should be named more cautiously. Nevertheless, since Dennett bases the force of his argumentation on the strength of his general examples as are to come, his criticism applies *both* to type as well as semantic physicalism.

¹⁹ Some commentators, e.g. Clark (Clark 2002: 67) and Soteriou (private correspondence) have attributed a 'whiff of Wittgenstein' with regard to this methodology to Dennett, and Dennett himself informs the reader to have been influenced by Wittgenstein during his philosophy studies (Dennett 1996). On the other hand many critics can be found, especially among the philosophers which are called 'Wittgensteinians', which are usually very eager to *deny* the any serious relation between Wittgenstein and Dennett. Though I myself do see a connection between these two philosophers, I hold the relation to be rather vague and therefore a deeper discussion of it not necessary.

Dennett is therefore more in line with the historical answer to identity theories, i.e. functionalism. Turing machine functionalism, i.e. the claim that for every mental predicate there is some predicate in a physically neutral language designed to specify abstract functions and functional relations is an example. To put the claim in different words, according to Turing machine functionalism, for 'X to believe that p' is equal to 'X realizing some Turing machine k in logical state A' (Dennett 1978b: xvi). This brand of functionalism clearly contradicts crude identity theories by allowing for multiple realization of a functional state. Dennett's position also allows for multiple realizability, so that in this point his position is again functionalistic.

But though functionalism is liberal, for Dennett it is not liberal *enough*. Dennett holds it to be not plausible that all believers that p could be described as Turing machines that are in the same logical state whenever they believe that p. As Dennett notes,

“There is no more reason to believe that you and I “have the same program” in *any* relaxed and abstract sense, considering the differences in our nature and nurture, that that our brains have identical physico-chemical descriptions. (Dennett, 1978b: xvi).”

Dennett illustrates his point by the example of three persons all believing one thing, e.g. that a Frenchman died last Saturday. Since presumably the three persons can have got to know about the event and have thought about the event in very different ways, there seems to be no reason to believe that there will have any striking physical or functional property in common simpliciter.

On these grounds Dennett argues that the sentence 'X believes that p' can best be understood by being translated into 'x can be predicatively attributed the belief that p'. Thereby Dennett purports to rebut the project of micro-reduction of intentional states as mistaken and proposes instead a project of *legitimizing* Turing machine talk by providing it with rules of attribution and exhibiting its predictive power, in other words, the adoption of the intentional stance. Dennett therefore names his position *type intentionalism*, i.e. the position that

“every mental event is some functional, physical event or other, and the types are not captured by any reductionist language, but by the regimentalization of the very terms we *ordinarily* use- we explain *what beliefs are* by systematizing the notion of a believing-system...(Dennett 1978b, xix).

To put the discussion in a nutshell, I have shown how Dennett rebuts micro-reductive approaches including functionalism. Though Dennett's position does include functionalistic elements, it is not best described as functionalism. I propose to follow this line further in Chapter 6.2.1 and 6.2.2 by comparing Dennett to Fodor due to following reasons. I would like to use the discussion first to end the explication of Dennett's negative task and second *at the same time* show that it is *exactly these differences* between Dennett's position and Fodor's position which are the motivation for Dennett to turn to macro-reduction in the form he does. Of course it wasn't only or primarily Fodor who shaped Dennett's further work, but I regard his criticism to be the best foil available against which Dennett can be described. In other words, the aim is to show that Dennett's critique of micro-reductive approaches *is responsible* for his turning to alternative approaches regarding topics like language, representation.

But there is more to Dennett's position. Dennett does not simply negate reductive micro-approaches and deduces the correctness of macro-reduction thereof. Instead, he asks the reader to conceptualize the whole problem of reduction *differently*. His argument is not that Fodor is wrong simpliciter, but that the idea which Fodor and others have of a successful solution reduction is ill-posed. He holds that belief in *any* kind of micro-reductive theory is misconceived and that the whole project of reduction must be re-conceptualized differently. One of the crucial points in Dennett's re-conceptualization of the task is his take on causality which will be dealt with in chapter 6.2.3.

6.2.Fodor's Criticism of Dennett and Dennett's Criticism of Fodor

6.2.1. Fodor's Position

Fodor's representational theory of mind can be looked at as an extension of the purely metaphysical doctrine of functionalism to a theory of the actual functional structure in question (Elton 2003: 21). The representational theory of mind holds that what is necessary to enable thought in a system is its possession of core elements and mechanisms explicated by equivocation of cognition with complex computer programs. A representation is hereby a type consisting of a set of tokens of physical objects that have causal powers and can *therefore* be accommodated by physicalism or naturalism easily. Representations are said to exhibit internal properties that are responsible for its intentionalistic character, for its aboutness and function as stand-ins for objects in the real world. Thinking, in effect, is the shuffling and manipulation of these representations. Representations in the mind can represent objects in the world, but also sentences like "Snow is white". A representation of the sentence "Snow is white" in one's head would in some way mirror the logical connection between the parts of

this sentence, i.e. words or concepts, in the structure of the representations. These representations are just physical objects, but in a machine equipped with representations and rules for shuffling representations one could, akin to computers, construct a symbol manipulating machine. This machine is just sensitive to the syntax of the representations, but in such a way that it mirrors the semantic properties of the representation in their manipulation. According to Fodor, what makes it true that one is a thinking agent is the possession of the relevant kind of representational structure and representation shuffling rules that are explicable as following causal laws on the level of syntax. What makes it true that an agent has a thought or belief is that the agent has a representation corresponding to the thought or belief in question in his head, e.g. qua a causal theory of reference (Fodor 1987).

According to Fodor these internal representations are arranged in a Language of Thought mirroring the structure of sentences. In its deep nature, the representational theory of mind is Cartesian in the sense that what makes it true that an agent holds a belief is not his overt behaviour, but rather that he does possess a representation in his head that mirrors the belief in question. As Fodor has the project to vindicate folk psychology and the notion of belief, he claims that for an agent to be in the possession of a belief the agent must be in the possession of a discrete, syntactic item in his head.

Fodor's representational theory of mind and attempt to reconcile intentionality in terms of aboutness and of-ness with naturalism is obviously contradictory with Dennett's position and Fodor himself is known to be one of Dennett's most vigorous critics. What exactly is it, then, that makes Fodor criticise Dennett's position so strongly? Fodor claims that Dennett's attempt to vindicate the folk psychological notion of belief is not *strong enough*, it is not *real* enough in a sense that Fodor believes to be crucial. According to Fodor for an agent to be in the possession of a belief is to be in the possession of a discrete item in the head that mirrors the belief.

6.2.2. Dennett, Functionalism and a Critique of Fodor

Dennett rebuts any kind of such an argumentation by arguing that since it is illusory to claim that any two systems are in general in any physical or functional way equivalent, they surely also do not have the same Language of Thought.

Fodor in turn is worried that a holistic interpretation of the concept of belief will make it impossible to individuate beliefs at all. According to Fodor, when two systems in question can be said to hold the same belief without having any relevant functional or physical property in common, belief individuation becomes impossible (Fodor 1992).

Dennett on the other hand believes Fodor's reasons to rebut holistic individuation of belief misguided by Fodor's use of *human language* as a model for the Language of Thought (Dennett 1987*d*, 1987*e*). Dennett instead favours a theory of belief that construes belief as rather primarily vague and without the use of internal criteria. This, though, is not the whole of the story.

Fodor charges Dennett to be an irrealist about propositional attitudes (Fodor 1985). In Fodor's eye for a propositional attitude to be real comes up to being a discrete item or representation in an agent's head. Nothing else will do the job, if it wants to count as real, since nothing else can play a role in a naturalistic world view that is governed by *causal laws*. If something is real, it is physical, and therefore subject to causal laws; if it plays any role at all, it does so by exhibiting causal efficacy. If beliefs are to be 'real', they have to be mirrored by representations in a subject's head, exhibiting causal efficacy qua the syntactic structure of the representation in question. Fodor goes on to argue that since Dennett construes intentional states not as representations in an agent's head but as ascribed predicates or at best as abstracta akin to centres of gravity, his position cannot be taken to be a realist position. Dennett is diagnosed either to fail to accommodate intentionality in a naturalistic worldview.

As I have tried to show in Chapters 2, 3 and 4, Dennett's early instrumentalist position seems to be especially prone to this criticism, but also that Dennett's later abstracta and illata distinction can be looked on as an attempt to reconcile intentional states with the kind of 'realist' assumption on causality that Fodor makes.

It has canonically both been argued that Dennett's position is either more realistic and nearer to the Fodorian sense of 'real' as it seems, or that Dennett's concept of real is quite different from Fodor's. My interpretation of Dennett *differs* from canonical interpretations. I believe there to be in both a grain of truth, and therefore set out to discuss both, starting with the latter possibility in order to arrive at a more balanced view.

6.2.3. Realism and Causation

As has been noted, the question of realism boils down to a question of causation. If an entity is supposedly 'real', it is made of purely physical stuff and obeys causal laws and vice versa. Fodor's concept of reality is exactly this kind of concept and the concept of causality the one used in the natural sciences and especially physics. On such a view beliefs can just be said to be real and cause actions if they cause these actions in the sense of causation as used by physics. Dennett rebuts such a concept of causation as too narrow and limited. As he puts it in an example, if a sign "Free lunch" causes people to visit a restaurant in order to get a free lunch, this is causality enough for him.

On this interpretation, Dennett's naturalism is quite different from Fodor's, since Dennett allows for a broad concept of causality whereas Fodor insists on the notion of causality as explicated in physics. Taking this view Dennett is not trying to do justice to a Fodorian type of causality, but to put forward an *alternative conception* of what we call 'real' or 'causality'. Dennett is hereby interpreted as stating that beliefs can be real and causally efficacious in the way presented and in the same way 'real' as e.g. centres of gravity. Moreover, Dennett can never be said to be a realist about the propositional attitudes as Fodor is, since normativity as a further characteristic property of belief on Dennett's account makes such a move impossible.

Dennett's position, if it were to be taken without the claim of beliefs being entities in normative space, and neglecting Dennett's criticism of functionalism as too narrowly construed, e.g. by putting forward a refined version of functionalism, could be taken to be a reductive theory. Though we are not to expect a simpliciter correspondence between beliefs and entities in the head of the agent, a *multiple realization* story could be told about revised beliefs and illata realizing these. Dennett's position, though, makes a claim of beliefs not only to be causally efficacious, but also to be *reasons* and therefore entities in *normative space*. Since purportedly normativity cannot be reduced to the interaction of causal interaction of physical entities, beliefs in Dennett's sense can never be as real as Fodor wants them to be. On this view Dennett is said not just not to try to do justice to a Fodorian concept of intentional states, but from the very beginning not to be able to do so on conceptual grounds. What Dennett does is to put forward an alternative concept of naturalism, causality and intentionality.

Up to this point I do not want to criticise the alternative, or better most common interpretation of Dennett, but I would like to oppose claims that are thought to follow from this interpretation (e.g. Elton 2003). What is often argued for, and especially on grounds of Dennett's account of causality, is that the intentional stance is in a particular way *independent*. It is the idea that looking at an intentional system one is able to switch from one stance to the other, always only changing the way of interpretation. From the intentional stance the system in question is evaluated according to normative considerations, whereas on the design and physical stance it is looked at as subject to causal laws as used in physics. Since one can 'flip' from one stance to another and still look at the same physical system in question, the intentional stance is no ontological danger to physicalism. This is true, but it is not the *whole* story that can be told about Dennett's account and that I take Dennett to tell,

since I believe the connection between the intentional and the lower stances to be more intricate than what the flipping account would like to make us believe.

It has to be observed in this regard that on the one hand Dennett stresses the independence of the stances, on the other hand he does *not*. He comments on the intentional stance to be nothing more than the design stance plus the assumption of optimal design (Dennett 1987c: 73) and he has an intricate story to tell about the relationship between *relata* and *abstracta* as already discussed in Chapter 2, 3 and 4. Furthermore, Dennett writes a lot about what can be told about the design stance and most of his recent work does explicate evolutionary stories of what kind of design could there be that underpins the realization of intentional system²⁰. Moreover, Dennett's *illata* and *abstracta* theory does try to create a connection between causality on the level of belief and causality on the level of subpersonal *illata*. I regard it to be important to stress this part of Dennett's account, since without it Dennett's philosophical position may seem weaker and more incoherent than it actually is.

First, I regard it as the best interpretation on the grounds given above to see Dennett not as suspending the notion of physical causation, but to draw our attention to the fact that the notion of causation should be conceptualized broader than before, since naturalism and substance monism can accommodate a broad notion of causation.

Second, Dennett's account of the relation between beliefs on the intentional level and content in the sub-intentional realm, if Dennett's epistemological story of content fixation at the subpersonal level and intentionality's role therein is neglected, seems to be as weak as it was criticised to be during Dennett's early instrumentalist phase. As I have argued following Dennett, the relation between the notion of intentionality and the notion of content on the subpersonal level is one of necessity when it comes to the discovery of content on the subpersonal level. If one leaves this part of Dennett's account out, his concept of intentionality may seem less 'real' than it really is.

Third, and in accord with earlier points, Dennett's philosophical program does rely on persuasion of the reader by examples as a form of conceptual clearing of the notions of belief and intentionality. Fodor can on this account to be interpreted as putting emphasis on the 'reality' of intentional states with the concept of intentionality being foremost the concept of *of-ness* and *aboutness*, but neglecting the normative dimension of beliefs as reasons. Attending to just the properties of the concept of intentionality may seem to make Fodor's account the stronger position as it comes to the reconciliation with naturalism.

²⁰ Chapter 7 and 8 of this essay can be looked at as a formulation and discussion of this thesis.

Dennett on the other hand can be seen as putting emphasis on the *primarily* predictive power of intentionality and *secondarily* on the thereby possible gains of explicability. His naturalistic account is meant also to do justice to the property of normativity as inherent in the concept of rationality, so that though his ‘reality’ of intentional states is weaker than Fodor’s- it is weaker because it also tries to accommodate normativity. Fodor for himself would claim that his account of causal reference does justice to normativity, but Dennett would rebut such a move on grounds that if Fodor does bring in normativity into the picture, he will not be able to bypass the considerations of radical interpretation and therefore end up, in the end, somewhere in Dennettian regions. This point of controversy is a major point of philosophical debate and its explication may require a thesis of its own, so that I cannot pursue the issue further at this point.

6.2.4. Wrap-up

The stage for the explication of Dennett’s newest philosophy is hereby set. I have explicated Dennett’s take on the relation between beliefs and physical states *both* in its positive form, i.e. pattern theory, intentional system theory and subpersonal cognitive psychology (Chapter 4) *and* in its negative form (Chapter 6.1-6.2), i.e. Dennett’s criticism on micro-reductive approaches. It has emerged that Dennett rebuts any kind of micro-reductive approach offered in the field. I have also shown that Dennett’s status as ‘realist’ depends strongly on his unique conceptualization of naturalism and offered an alternative reading to the canonical one of Dennett’s take on the relation of the stances. I find it better compatible with Dennett’s newest writings. These points together are responsible for Dennett’s turn to macro-reduction.

6.3. Macro-reduction

As has become visible by now, Dennett holds that the explanation of the nature of intentional states has to be elucidated in a different sense than through micro-reduction. His proposal is to turn to macro-reduction in general and evolutionary theory particularly. Already his early works (Dennett 1969, 1971) contain allusions to the role of evolution as the last explanatory level at which intentionality would come to be explained. In “Three Kinds of Intentional Psychology” (Dennett 1981a) Dennett shortly discusses Friedman’s suggestion to reverse the usual order of reduction by first looking for a theory of social activity and then try to reduce individual cognitive ability to social activity. Dennett acknowledges the important role of social activity, but also observes a clear relation between social activity and language

that in effect does not do justice, as it comes to macro-reduction, of non-language using intentional systems. Dennett therefore notes that

“The macro level *up* to which we should relate micro processes in the brain in order to understand them as psychological is more broadly the level of the organism-environmental interaction, development, and evolution. That level includes social interaction as a particularly important part... (Dennett 1981a: 65)”

Dennett hence notices semantic properties not just to be relational, but ‘*super-relational*’,

“...[f]or the relation a particular vehicle of content, or token, must bear in order to have content is not just a relation it bears to other similar things...but a relation between the tokens and the whole life- and counterfactual life- of the organism it serves *and* that organism’s requirement for survival *and* its evolutionary ancestry (Dennett 1981a: 65).”

Dennett’s construal of dependence of content is very rich. It includes in general organism-environmental interaction, i.e. the synchronic position of an intentional system in the world, wherever it may be, the development, i.e. biography of every single intentional system in question, and its phylogenetic evolution. If that many factors do fix the content of brain processes, the question can be raised which factor has to be applied when and if two factors can be thought to fix content differently, which one trumps the other? Dennett’s possible answer to the question can be deduced from his general philosophical position and a multitude of scattered remarks throughout his work.

First, it is to be expected that if any factor does trump the other, synchronic factors will trump diachronic ones. As Dennett notes, human artefacts like thermostats and lanterns can be interpreted as intentional systems without possessing a history of natural selection. Priority of explanation is to be given in how intentionality as *involvement in the world* comes about, i.e. be described by intentional idiom purely on the basis of their behaviour and the assumption of rationality.

Second, in the case of two factors fixing content to the contrary Dennett most notably comments that maybe there simply is no matter to the fact whether this or that content is in question. The ground for such an argumentation, of course, is Quine’s radical interpretation thesis.

If we accept Dennett’s position and argumentation with regard to macro-reduction of intentional states, we find ourselves before an open field open to a multitude of possible ways of fleshing out. To Dennett this is not just a possibility, but a *necessity*. Having rebutted Fodor’s realism and language of thought Dennett has taken the burden to offer another

explication of how intentional states could be realized in a system, if not through mental words and sentences. Natural candidates for endorsement are connectionism and ‘new cognitive science’, relating the explanation of the mind less to digital computers, but rather to embodied embedded cognition, or evolutionary psychology. It is the latter of these two that Dennett explicates in many recent publications and books (e.g. Dennett, 1995, 1996*b*). Though Dennett does in no way judge explicit, language-like content bearing states to be impossible in an intentional system, he draws attention to other ways how information, knowledge or content can be ‘in’ a system (e.g. tacitly (Dennett, 1982*c*)) and according to Dennett do play the primary explanatory role in any account of content fixation. Dennett’s account of language-like content bearing states does play a huge role in Dennett’s explication of why and how it is special to be human, but it is a very special kind of possessing content bearing states according to Dennett and no basic faculty.

Furthermore, Dennett tries to do justice to evolutionary thinking by construing the phylogentic development of mental states as gradual and the mind not as a general problem solver, but as a bag of tricks (Dennett 1981*a*), as a set of tools (Dennett 1996*b*) that has developed from and on top of evolutionary older structures. As Dennett’s philosophy is not limited to content fixating properties being internal to the system, his account quite readily opens up towards not only content externalism, but also ‘mind externalism’, a gloss for Dennett’s position that intentional states are not only in a subject’s head, but also out there in the world, involving the world, in creating real patterns. It is in this boiling mixture of evolution, patterns and content that Dennett’s quite radical account of the human mind as a set of tools does arise.

The complete topic of how and why tools do make sense as a model of the human mind cannot be discussed here, since the whole story incorporates Dennett’s special position on the nature of consciousness, the role of language in consciousness, Dennett’s ambitious adaptionistic project as started by Richard Dawkins and many other factors. What does lie in the limits of this paper is a short introduction and discussion of how Dennett’s newest account of the human mind as a set of tools does relate to his prior theory of intentionality and how it fits in with Dennett’s take on language and mental representations. These topics will be discussed in the next two chapters.

7. Representation, Language and Intentionality

Revisited

7.1. Introduction

A heavy burden of mandatory explanation lies on Dennett's shoulders (Chapter 6.3). His task is difficult because he has to furnish us with a unified and stringent account of intentionality that fits all intentional systems and *at the same time* take care that he does not leave out important properties that the intentionality of some, but not of all creatures possesses. He has to explain similarities and differences at the same time.

Most unsatisfied with Dennett's performance in this regard are his perennial critics, Fodor and Searle. Acknowledging even that Dennett's account of intentionality may in some way be interesting, they also stress that the intentionality of human beings is rather different from that of thermostats. The worry is partly the old one, but there is also a new part in it, i.e. that our intentionality is conscious, it is not free-floating, not the kind of intentionality of a robot. To acknowledge the force of this criticism I ask the reader to wait for an exposition of the explanation that Dennett wants to offer for the genesis of intentionality.

Moreover, it is not only the intentionality that differs, but also many other facts about human cognition. Human intelligence seems to be rather different from the intelligence of animals. Humans communicate, they use tools, they are immersed in human culture and advance technologically at a rapid rate. Dennett states that any kind of interesting explanation of intelligence or cognition would have to acknowledge the existence of the intentional stance, and granting this point, the question remains what it is that makes human beings such astonishing members of the set of intentional agents so that only these have developed all these fascinating capabilities. Dennett owes a developmental account of this observation.

Thirdly, and strongly related to the last two points, Dennett himself notes that beliefs and desires in a way are well explained on the intentional stance, but that at least some of '*our*' human kind of beliefs and desires are still rather different from the one's animals possess. Dennett notes and acknowledges from the very beginning of his work that some human beliefs and desires are in a different way intentional because they allow for linguistic, precise and propositional description, whereas animal beliefs do not. In 'How to change your mind' (Dennett 1978c) and 'True believers' (Dennett 1981b) Dennett as a result makes a distinction between beliefs and *opinions*, i.e. linguistically infected, relatively sophisticated

cognitive states, roughly the states of betting on the truth of a particular sentence and holds these, at first without explanation to be ‘only *derivatively* and *problematically* contentful states of language users’ (Dennett 1978*b*).

Dennett observes that therefore in the realm of belief language plays a double role—first, it enables one to formulate highly specific, propositionally specifiable desires and second, forces commitment to more stringent desires than that would be endorsed without language infections (Dennett 1981*b*: 20-21). The ‘social force’ producing these linguistic results in Dennett earlier accounts takes centre stage in explanations of the difference between human and animal belief. Dennett adds that though beliefs in linguistic form ‘seem’ to exist, they are not a good model for the whole of the mind, but also leaves the reader without an explanation of *how* and *why* then language does play its crucial role. Such an explanation, as Dennett notes himself in another context, will have to be one of macro-reduction (Chapter 6.3). This macro-reduction will prove to be a rather highly complicated and intricate matter given Dennett’s limits to what a good philosophy of mind is supposed to look like.

Dennett’s newest books, starting with ‘Consciousness explained’ (Dennett, 1991), Darwin’s dangerous idea (1995), ‘Kinds of minds’ (Dennett 1996*b*) and a massive amount of essays do address exactly these questions and the implications of the answers. Dennett tries to furnish us with a complete explanation of how human intelligence and culture could come about without intrinsic, essentialist concepts of intentionality, without a deliberate, intentional creator, merely based on evolution and starting with macromolecules. To make his naturalistic account plausible Dennett has to explain how complex intentional systems can come into being, how they can be constructed of and derived from mere robotic or macromolecular matter.

Biological evolution hereby is not the only route that Dennett takes and even can take, observing the massive impact of culture on the human mind. Dennett therefore puts forward a highly ambitious concept of *cultural evolution* encompassing a concept of *language* and a concept of *tools*. The development of both language and tools has thereby to abstain from any kind of deliberate design as a foundation. Deliberate design may be the outcome, but not its origin if ‘the question is not to be begged’. Dennett’s concepts of the ‘tower of generate and test’ (Dennett 1976), the idea of ‘found objects’ (Dennett 1996*b*) and the adoption of the concept of *memes* as introduced by Dawkins (Dawkins 1976, 1982) are thus Dennett’s main tools.

The influence of language is explained by what Dennett desires to call ‘representations’. Language is hereby given a primal role in human cognition and Dennett’s

account fills out the explanatory gaps in his account that he hinted at by stating that ‘some version of [RTM] will prove correct (Dennett, 1981*b* : 34)’.

The influence of tools is stipulated by Dennett to be responsible for a major boost for the development of human intelligence. Though Dennett is not the inventor of the concept of tools, his take on this issue is unique. For Dennett tools are not only laptops or calculators, but also the shopping list in one’s mind or the rehearsed and automatized skill of addition in one’s head. The most important human tool is a mind-tool, i.e. *human language*.

Language therefore seems to play at least a *three-fold role* in Dennett’s account. First, it is the major and fundamental *representational scheme* available to human thought, second, it is a *mind tool* and third, it is a *meme*. The conceptualization of the interrelation of all these concepts is what makes Dennett’s account so unique, not the ideas themselves. Unluckily it also makes his account highly complex.

To examine Dennett’s account of biological and cultural evolution of intentionality I propose to take the following way. First, I would like to discuss Dennett’s account of the *genesis* of intentionality (Chapter 7.2). I will defend the hypothesis that this account changes Dennett’s concept of intentionality from a predictive to a teleological position. Next I would like to discuss the status of the concept of representation in Dennett’s philosophy (Chapter 7.3.1 and 7.3.2) in order to get clear in how far it solves the problems posed in this introduction and elucidate its relation to the role of language (Chapter 7.3.3). Building up on these findings I will be able to tackle Dennett’s general take on mind-tools in Chapter 8.

7.2. The Genesis of Intentionality and the Conflation of the Stances

Dennett’s stand on the issue of the genesis of intentionality is to provide an account of the origin of intentionality as a way to get informed about the mind without the assumptions of a creator (Dennett, 1995, X). Intentionality must come into being without original intention, and this process must be shown to be able to create *human intentionality* as well. It is an evolutionary story that Dennett turns to in order to provide convincing evidence for his position.

By doing so Dennett’s new concept of intentionality changes considerable in so far as that it *conflates* the design with the intentional stance. This is more than what I have argued for in Chapter 5. In the time of Dennett’s pattern theory (Chapter 5) Dennett kept the two stances apart by motivating a normative dimension in the intentional stance. I hypothesize that in his newest writings the kind of normativity, if any, that can be found on the intentional stance is the same as the one that can be found on the design stance.

How then did, according to Dennett, human intentionality derive from mere aboutness or intentionality of macro-molecules? Dennett provides a framework for the development of differently powerful and complex design options for brains, representing the increasing power of systems to do what Dennett takes to be most important in the generation of intelligent behaviour, i.e. to see farther away in space and time. Dennett's newer theory thereby is derived from his behaviouristic concept of the tower of generate-and-test, but a full discussion of the development of Dennett's take on these special issues is beyond the scope of this essay (Dennett 1976, 1996b).

Dennett takes *agency*, i.e. the possession of enough complexity to perform actions instead of just having effects (Dennett, 1996b: 21), to be a precondition for any possession of intentionality. Though Dennett seems to put emphasis on *actions*, it is clear that Dennett also binds the ability to discriminate actions from effects to enough 'complexity' of the system and thereby stresses the importance of the system's *design* that has to be furnished in such a way as to behave in a complex fashion²¹.

Action is hereby tied in the first place not to our kind of deliberate and conscious action, but merely to the exhibition of systematic behaviour. Replicating molecules as robots, i.e. 'impersonal, robotic, mindless entities (Dennett, 1996b: 24)' on this account already count as agents and are the (constitutive) basis for any other kind of more complex intentional system.

Therefore the first level of Dennett's framework is inhabited by entities exhibiting agency and being selected by natural selection. Nothing more is necessary to inhabit this first level. Dennett terms its inhabitants *Darwinian creatures*.

The next level is occupied with *Skinnerian creatures*, i.e. creatures that exhibiting phenotypic plasticity through the ability to adjust elements of their design by events that occur during the biography of the creature unconsciously and unwillingly. These creatures are said to possess hard-wired reinforcers that happen to favour what Dennett calls 'smart moves', i.e. behaviour that is selectively adaptive. They generate a variety of actions and try out until they find out what works and what doesn't. Dennett thereby takes up Skinner's famous remark that 'operant conditioning is an extension of evolution' (Skinner, 1953 in Dennett 1996b: 83), where the environment plays a blind but selective role in shaping the behaviour of the creature. One shot learning is not possible on this level and a mistake may be fatal to the creature.

²¹ At this point the reader may be reminded of my interpretation of Dennett favouring to view rationality in terms of a system's design rather than in terms of a system's actions Chapter 5.2).

On the next level, *Popperian creatures* are capable of one-shot learning by pre-selecting among possible behaviours. This design enhancement permits, as Popper put it, 'our hypotheses to die in our stead'. Popperian creatures have an evolutionary advantage because they make better than chance level first moves as compared to Popperian creatures. To allow for pre-selection Dennett proposes the creatures to possess a filter, an inner environment that in some aspects mirrors the environment and possesses a lot of information about the outer environment in a way that allows trying out moves in it. If the inner environment mirrors the outer environment in the relevant ways and the information in it useful, may it be inherited or acquired via the senses, the best move given the trial runs can be selected. The special nature of this inner environment will be discussed later in the essay (Chapter 7.3.1).

The next step, probably Dennett's most ambitious conjecture, are *Gregorian creatures*, i.e. Popperian creatures whose inner environments are informed by designed portions of the outer world (Dennett 1996b: 99). The use of informed portions of the environment is said not only to be the result of intelligence, but also to *endow* the user with intelligence. Dennett terms these designed items of the environment *tools* and observes that the advance of tool use in human history has accompanied and is still responsible for a major increase in intelligence. Among these tools, Dennett recognizes one kind of tools to be most important, namely mind tools, and in that category, following Gregory, words as the most important ones (Dennett 1996b: 100). Words and other mind tools themselves then allow humans to construct even more subtle generators and testers, i.e. better informed inner environments by exploiting wisdom embodied in the mind tool that others have invented, improved and transmitted. As Dennett puts it in a phrase, Gregorian creatures start not only to think what to do, but to think about what they shall think next. By the use of tools humans as informed Gregorian creatures gain higher tracking abilities in space and time and reflective power. Language as a tool and fundamental biological talents are supposed to permit the human mind to extract these mind tools from the (social) environment in which they reside (Dennett 1996b: 117).

Giving this hierarchy Dennett goes on to argue how the different levels fit together, i.e. how one level could develop out of the other. In accord with the observation that evolution cannot furnish new design in any way except by slightly changing old design (Chapter 5), Dennett stipulates that *development* from a lower form to a higher one can be conceived as the higher one *being composed of* the design of the lower. Though humans as exemplary inhabitants of the highest stance are not robots themselves, they are supposed to be composed of robots (Dennett, 1996b: 22).

Dennett goes on to argue that since humans in this sense are composed of robots, the same considerations concerning the application of the intentional stance apply to all levels. Since agents are information-modulated and goal seeking systems, they are apt as candidates for prediction from the intentional stance *as a subclass of the design stance* assuming that whatever the design of the entity in question may be, it will behave rationally and act for good reasons (Dennett, 1996b: 31).

Dennett thereby observes, in contradiction to this observation and to his earlier writings, that the intentional stance as applied to e.g. macromolecules is a predictive strategy treating the entity *only as if* it were an intentional agent who governs its choice of action by consideration of its beliefs and desires (Dennett, 1996b: 27). It cannot be observed, though, that Dennett has changed his mind and attributes rationality and intentionality only to one kind of rational agent, namely us, so that the status of the *as if* is not a difference between real intentionality and false intentionality, but one between the degree of *similarity* between some intentional system and *our* mind. What Dennett *beyond that* has in mind is highly unclear. On the one hand Dennett reuses his account of the intentional system approach, on the other he states that ‘Something exhibits intentionality if its competence is in some way about something...Intentionality in the philosophical sense is just aboutness (Dennett 1996b: 35)’. Furthermore and in connection with the attribution of goals and the possession of information, Dennett observes that if an entity is designed, it has some end that is good for it and thereby introduces the crucial notion of *function* (Dennett 1996b: 32)

Hereby Dennett’s position shifts from an interpretative to a teleological account, since nothing of what he says about intentionality could not already be said merely on the adoption of the design stance. The intentional stance still exists as the level on which design is unimportant, but only on the assumption that there *must be some design, whatever it is*, that works well. This obviously is not enough to keep the stances apart.

Furthermore, as a negative reaction to Searle’s distinction between derived and original intentionality Dennett proposes the following account. Dennett agrees that in some sense designed artefacts derive their intentionality from their originators, i.e. their creators, but states that this argument does not carry over to mental states. Mental states do not possess intentionality due to intrinsic intentionality neither as is purportedly provided by mental languages nor by consciousness, but by being itself an artefact. Mental states get their intentionality from their ‘particular position in the ongoing economy of the brain’s internal activities and their role in governing your body’s complex activities in the real, surrounding

world' (Dennett 1996b: 52). Brains as minds²² do get their intentionality thereby again from their role in the ongoing economy of the larger system of what it is part, and that is evolution by natural selection. As Dennett puts it, "derived intentionality can be derived from derived intentionality and there is no place or necessity for intrinsic intentionality (Dennett 1996b: 54)."

Therefore, human intentionality is derived and build from as-if, derived intentionality (Dennett 1996b: 55). Since derived as-if intentionality is supposed to be completely explainable by evolutionary thinking concerning design, nothing substantial distinguishes the intentional stance from the design stance anymore. The two stances collide.

Consequently one may wonder where the assumption of 'rational' design has gone (Chapter 5.1), and what it is over and above what evolution has given us that makes a design rational²³. One might be tempted to view rationality and the intentional stance as something 'cognitively high-level', as maybe only available to complex agents, but it is hard to make sense of such an interpretation on Dennett's account. As we already know, Dennett ascribes beliefs and desires to thermostats. Furthermore, he describes macro-molecules to have needs (Dennett, 1996b: 57), and plants as possessing 'highly distributed *decision making* (Dennett, 1996b: 65). He also recognises *reasons* being ascribable to macromolecules, though these are not reason generators or 'understanders'. Reasons that a macromolecule may possess are therefore to be described as free-floating rationales, as patterns (Dennett 1996b: 60). We do not tend to see these reasons and overlook the rational behaviour of a species over time due to what Dennett calls time-scale chauvinism, though species can be said to be sensitive to changes over time and respond to the changes in rational ways. Dennett treatment of the concept of understanding is thereby highly confusing. We are told that macro-molecules have needs, act rationally and have beliefs, but that they do *not understand*, and it is hard to reconcile the ascription of all these properties without wanting to attribute understanding to it. As Dennett notes, to be an understander is to have a mind (Dennett, 1996b), so one may follow from this that whatever there is that is not a reason generator is not an understander,

²² Dennett's sloppy equivocation of 'minds' with 'brains' may be rather disturbing to a reader who knows Dennett as a critic of any kind of identity theory (Chapter 6.2.2). And even if one assumes Dennett to have changed his mind and suddenly to allow for identity, such a claim does and cannot make sense in connection with Dennett's position on mind externalism and external cognition (Chapter 6.3, 7, 8). It furthermore makes little sense, since brains are nor per se the objects of natural selection. To make sense out of Dennett's equivocation I first propose to view it as Dennett's way to by-pass his real position on the relation of mind and matter for the sake of the presented argument and second to exchange 'brain' by 'whatever it is to what the mind is supposed to be micro-or macro-reduced'. Therefore the argument becomes that whatever it may be to what the mind is reduced it is that which gives the mind its intentionality by being the object of natural selection

²³ The reader may be reminded of the prior discussion of rationality in Chapter 5 which shall not be repeated here.

and therefore does not have a mind. The conclusion that one may have reasons and hence be rational, but not possess a mind, strikes one as highly counterintuitive.

To make sense out of this muddle I propose the following reading of Dennett. In its late work Dennett has for the most dropped the assumption of rationality as optimal or as cognitive approval. All there is to be rational is a pattern that is visible and that in a very weak sense is used or followed by the entity in question plus environmental interaction. Dennett dissolves the concept of rationality in evolution and introduces a teleological, Millikan-like (Millikan 1984) concept of function (Dennett 1996a) instead which is supposed to provide for the normative dimension of function and to be constitutive for any normativity that is exhibited by the use of intentionalistic idiom by linguistic creatures. What Dennett does keep is the idea of there not being one decisive function given by evolution intrinsically, but the need to interpret it under the constraints of radical interpretation. Hence, as already hinted at in the beginning of the thesis, the intentional stance and radical translation per se may be seen as *epistemological methodology* to gain access to the entities that possess intentionality, but once we get there, it is the entities that are ontologically fundamental.

In sum I take Dennett to collapse the design with the intentional stance. Dennett tries to keep the two apart by describing the assumption made by the intentional stance as dependent on behaving rationally and due to reasons. This could work if reasons and rationality were something possessed by humans alone, but Dennett's account of rationality and reasons does allow reasons already on the level of macromolecules. If we can talk about reasons and about rationality already on such a low level, there is no possibility to secure any special standing to the intentional stance.

Given this result several points emerge. By conflating the stances Dennett reduces the possibilities of explanation of any kind of intentional phenomenon to exactly one: evolution. Therefore the view of humans as being composed of robots is not only an explanatory opportunity, but a *necessity*. This being so, Dennett has to offer an evolutionary account of whatever concept he uses to explain intentionality, both in its simple and *complex*, i.e. human form. In the case of Darwinian and by extension Skinnerian creatures, natural evolution is enough. In the case of Popperian creatures Dennett makes use of the concept of *representation*. Dennett therefore has to offer an evolutionary explanation of representation. In the case of Gregorian creatures, the situation gets even worse. Here Dennett also has to furnish us with an evolutionary explanation of tools and language. Pure natural selection is thereby not enough, explaining Dennett's turn towards cultural evolution and meme theory.

As according to Dennett the concept of representation appears before the concepts of language in the hierarchy, I propose to discuss it first before discussing language and tools.

7.3. Language and Representation

7.3.1. The Beginnings of Representation

Dennett's take on mental representation and language shows a strong development during Dennett's work. The basis for understanding what Dennett does *not* mean by mental representation has already been laid in Chapter 6.2, i.e. that mental representations must not play any essentialist or fundamental explicatory role. Dennett's *positive* take on the notion of representation has been in his account from the very beginning which is highly confusing and contradictory.

As a reason to believe in mental representations Dennett cites that some kind of generative, indefinitely expandable mechanism of representation must be present to solve the problem of *combinatorial explosion* of thought, i.e. explain the richness and fullness of possibility of content, in which case human language seems to be the only possible model.

At the same time Dennett's first idea of mental representation is highly influenced by his ideas of individuating mental states holistically and of the concept of representation to be properly used in the following exemplary case. Imagining a complex intentional system in a rich behavioural surrounding Dennett observes a two-way constraint on growing specificity between the intentional system and the environment (Dennett, 1981*b*: 31-32). If the system is not being fixed in one state and the environment changes, then the subject's sensory attachments will be sensitive and discriminate enough to respond to the changes appropriately. One may say that due to the *tight relationship* between the organization of the system and the environment the organism *mirrors* the environment by being sensitive to and tracking the changes, or that there is a *representation* of the environment in the system. It has to be noted that here an internal item is made a representation only by its role in regulating the behaviour of an intentional system and that according to Dennett, representation is a *gradual concept* that gains usability in use with growing ability of 'mirroring' the environment in a complex fashion.

Any kind of representational theory therefore must gain its plausibility from this property of systems and cannot be taken for granted. Dennett considers representations as explananda of some fundamental theory of the workings of the mind, not as explanans (Dennett 1981*a*).

Dennett's position obviously poses problems to understand what is meant by mental representation at all. His idea of representation being a gradual concept allows him to use the

word rather carelessly as e.g. in reporting and acknowledging Sellars (Sellars 1974) to have recognized the necessity of behavioural analysis of the semantic properties of inner representations (Dennett 1987*b*: 346) and stating that Bennett, as himself, sees the need to build an account of human belief on top of the representational resources of animals (Dennett 1987*b*: 348).

Such sloppy usage is more than even the most charitable reader is able to bear. Even if one acknowledges the gradual nature of the concept of representation as build on some kind of ‘mirroring relation’, it is hard to make sense of the claims that on the one hand the only plausibility to believe in mental representations comes from belief in the existence of a generative and indefinitely expandable mechanism of thought whose only model can be language and at the same time to talk about the *representational capacities of non-human animals* and sub-language systems, while denying animals the possession of language. On the one hand Dennett wants to introduce representations as ‘something like’ the Representational Theory of Mind’ on top of a more fundamental theory of mind (Dennett, 1987*b*: 348, 1981*b*: 34), on the other hand he uses the concept of representation, besides the mysterious remark of some solution to combinatorial explosion, as not very different from the concept of an intentional state. On the one hand representations, as intentional states, are *about* something and both are gradually attributable, but on the other hand Dennett holds intentionality to be more basic than representations. As he notes in ‘Kinds of Minds’ (Dennett 1996*b*), the intentionality of the lock and key variety as displayed already on the level of macromolecules is not representation, but the *basic design element* out of which nature has fashioned representational systems (Dennett 1996*b*: 35).

To make sense out of Dennett it is possible to put forward the following interpretation. Dennett wants representations to be more than intentionality, but in the end fashioned out of the same material or principle as intentionality is. Therefore, representations are intentionality plus *whatever is necessary* to explain combinatorial explosion as it can be witnessed in language and modelled like language. The questions that Dennett is in debt to answer are hence first, what exactly this ‘*whatever*’ is, second how representations can be thought to derive from basic intentionality and third how in the end the concept of representation relates to human intelligence.

I take Dennett’s paper ‘Styles of mental representation’ (Dennett 1982*c*) to be an attempt to clear up a part of the confusion of the concept of representation that prevails in his earlier work, though most of Dennett’s later work could by the uncharitable interpreter also be looked at as ‘blind shots in the darkness’ of the concept of the ‘whatever’.

7.3.2. Styles of Mental Representation

Dennett's 'Styles of mental Representation' (Dennett 1982c) is a locus classicus of Dennett's overall philosophical position as applied to the problem of the concept of mental representation. Following Ryle Dennett recognizes that human practices as conscious deliberate consideration of stated maxims, following of rules, deduction and inference depend on the agent's possession of full blown *talents*. In this context the use of the computer metaphor, according to Dennett, has been misleadingly used and a clearing up of what it is the case when computers represent is necessary to understand what is the case when human minds represent²⁴.

To explicate the workings of a computer, Dennett recognizes three ways in which information can be in a system: explicitly, implicitly and tacitly. Dennett defines information to be represented explicitly in a system iff there exists in the functionally relevant place in the system a physically structured object, a formula or tokening or string composed of members of a system of elements for which there is a semantic interpretation and a mechanism for reading or parsing the formula. In this sense, explicit representations are syntactic, structural elements made out of symbols that are made up according to composition rules and make complex meanings computable out of the meanings of their parts. According to Dennett, this is the kind of representation that the RTM has in mind. For information to be represented implicitly it is meant that it is logically implied by what is stored explicitly, e.g. if 'p implies q' and p are represented explicitly, then q is represented implicitly in the system. Dennett notices that though explicit and implicit representation by classical cognitive science are usually taken to be the main players in the explanation of the working of the human mind, this can in no way be so, since explicit information bearers are inert in isolation and become 'real' information bearers only in larger systems. It is not exactly clear whom Dennett wants to criticise as forgetting this point, since even Fodor does acknowledge this point freely. Dennett's idea hereby is better interpreted rather differently. As an argument, Dennett states that if only explicit representation were to be said to represent, then it would follow that information cannot be sent or held without being represented, which is highly counterintuitive. Therefore, the notion of representation has in some sense to be extended to apply to whatever bears information, i.e. has content or is about something.²⁵

²⁴ Of course this argumentation presumes that in some sense the computer metaphor *is* a good model of the human mind. The crux to understand Dennett is to see in how far he really *does* believe the computer metaphor to be a good model of the human mind.

²⁵ On first sight this conclusion seems to contradict my proposal to view Dennett's take on representation to be intentionality plus whatever is necessary to explain combinatorial explosion, since the reduction of the concept of representation to the concept of information seems to be rather better interpreted as simple alignment of the

Having glossed possessing a representation as bearing information, Dennett renews Ryle's well known statement that know-that depends on know-how, by aligning know-that with explicit representations and know-how with the agent's talents and skills. These talents and skills, which are not represented explicitly or implicitly, are according to Dennett represented tacitly. As a general argument to avoid general regress, Dennett observes that there must be a system beneath explicit representation and manipulation that merely has the know-how. If this were not so, the question 'How does the system know how to follow a rule?' would have to be answered by 'By following another rule' ad infinitum. But since there is information in the know-how, i.e. how to follow a rule without following another, this kind of know-how has to be called representation, i.e. tacit representation. To understand or to follow rules human beings must be able to understand or follow a rule without having to understand these. At the bottom level, the system must be stupid, without understanding, working brutally in terms of storage and manipulation. Know-how can thereby be merely tacit, or itself based on some internal rule-following process of narrower horizon and greater stupidity (Dennett 1976, 1978a).

Explicit representation can from this reason neither be looked at as explanatorily nor ontologically prior to tacit knowledge. Since explicit representation is inert and 'senseless' without tacit knowledge, it is explanatorily secondary, and since this is so, tacit representation must have come before explicit representation in the history of the development of the mind. Dennett offers an explanation of how explicit representation could have emerged out of tacit representation that fits in with his framework of the development of basic intentionality into intelligence (Chapter 7.2). One shall imagine simple creatures, i.e. animals made up of devices that obey rules without consulting any expression of them (Dennett 1982c: 222). Further one can imagine a whole set of these tacit rules being represented in an animal at the same time, but depending on the state of the system in turn depending on external cues, only one set of the rules tacitly represented is being followed at any time. This following of the system by this set of rules then can be seen as a representation for whatever it was that caused the set to be followed. The non-activated set of rules one may call transiently and tacitly

concept of representation with the concept of intentionality. To clear the situation I propose the following interpretation of Dennett's remarks on this issue before giving another interpretation at the end of this part of the chapter. Dennett's program is two fold. On the one hand he does want to stress the similarity between intentionality and representation, on the other he wants to stress the differences between the concepts. The same dichotomy can be witnessed in Dennett's discrimination between beliefs and opinions (Dennett, 1978c), while at the same time stressing that explicit representation has to be derived from skills in some way. One may feel that Dennett thereby shall better also divide up the terms he uses into two, since these evidently refer to different concepts. Though such a change could make Dennett more understandable in the sense of being precise, it would undermine his idea of unification of phenomena under the idea of gradual conceptualization. I propose to let Dennett do how he wants as long as one can infer what he is talking about.

represented. If one imagines such systems to become more complex, then there is no limited specificity as to what could be tacitly represented by the system. With a growing number of possible states that do not differ much, it becomes mandatory or at least more efficient not to switch from subsystem to subsystem, but to change one or more of the features of the current subsystem, that is changing states by editing and revising, not by discharging and replacing²⁶. This in turn requires systematicity, since the loci at which substitutions can be made have to have a fixed way of changing their function as a function of the identity of their substituents. Dennett notes that with growing ability to speak of systematicity in this case, the notion of *explicit* representation gets a grip. This view is in accord with Dennett's idea of representation mirroring the environment, where tacit representation simply mirrors, whereas explicit representation mirrors *systematically*. This ability to systematic change is also what Dennett believes to make explicit representation look like a language, as the kind of device that is needed to solve the problem of combinatorial explosion due to better resource management by systematic recombination.

To sum up, in 'Styles of Mental Representation' Dennett answers two of the three questions posed above. First, the relation between representations and intentionality is cleared up. Representations are not mere intentionality, but depending on their nature as explicit, implicit or even tacit, they are more. Tacit representations are intentionality *as exhibited in* skills or talents, explicit representations are intentionality plus skills and talents organized *systematically* in a unique way, and implicit representations are potential explicit representations that behave exactly like explicit representations if realized, activated and instantiated.

Second, as to how representations can derive from basic intentionality Dennett's answer is to hint on basic, tacit skills becoming systematized due to efficiency by natural selection, in accordance with Dennett's evolutionary position as portrayed in Chapter 7.2. Dennett's position is to speak of representations any time one can speak of information being somehow in the system, and of explicit representation as arising as an adaptation to manage complex behaviour by organizing and structuring mental states systematically. Therefore, in 'Styles of mental representation', the need of systematicity in order to explain the complexity of behaviour without having to posit an infinite number of possible states, i.e. devices that

²⁶ Dennett uses exactly the same terminology pairs (editing, revising/ discharging, replacing) in 'Consciousness Explained' in the explication of what makes content fixation in his multiple draft theory special (Dennett 1991: Chapter 5) and different from traditional models of content fixation. This observation points to Dennett's general conceptual methodology with regard to argumentation: the same metaphors are used repeatedly in different contexts and in different levels of explanation in order to convince the reader to adopt a unifying perspective on the mind.

follow some rule and are describable as a skill, but only a finite number of systematically changing states, has been the motivation for the postulation of explicit representation. Dennett's further work offers many more possible reasons and motivation of when and why to posit explicit representations in a system (Chapter 7.3.3).

The third question posed above, i.e. how the concept of representation relates to human intelligence is left unanswered. Recognizing language to be the main ingredient of human intelligence Dennett merely hints at the connection between explicit representation and language, may it be LoTH or language as a human communicative device. It is to Dennett's further thoughts about explicit representation and their relation to language that I would like to turn in order to answer the third question posed above.

7.3.3. Explicit Representation and Language

To understand better what Dennett believes to be explicit representation and what how he puts this concept to work, it is necessary to turn to some of the examples he gives to elucidate his position.

According to Dennett, animals, though thinking creatures, are not persons, since they are at most first order intentional systems and not, like us humans, second order intentional systems (Dennett 1976). They do not represent their intentional states explicitly. This makes animals, though being thinking creatures, possess a different kind of mind than we do (Dennett 1996*b*: 121). According to Dennett's view on empirical results in animal ethology, animals don't represent the minds of others and consult the inner model to anticipate the other's behaviour.²⁷

Dennett also notices a different and supporting need for explicit, manipulable representation. Following McFarland, Dennett states that the need for explicit, manipulable representation arises when the option of potentially cooperative, but still self-protective *communication* emerges, since in this case the agent faces a new behavioural task, i.e. explicitly communicating something about one's other behaviour (Dennett 1996*b*: 127).²⁸ In

²⁷ A discussion of this highly debated hypothesis cannot be given in this essay, since this topic requires extensive treatment of its own. The status of animal theory of mind is being debated under the headings of e.g. theory-theory and simulation theory and forms an own subfield in the philosophy of mind. A well known critic of the sceptic position, who in many other aspects agrees with Dennett on the status of 'animal psychology', is e.g. Susan Hurley (Hurley 2003). The interested reader may consult any of the immense number of essays on this topic.

²⁸ This problem of explicitness in turn is, according to McFarland and in great harmony with Dennett's own position of the emergence of the self (Dennett 1991: Chapter 5, 1992), solved by confabulation, i.e. the agent comes to discriminate and label his own tendencies as if they were governed by explicitly represented goals and plans. Once such representations of intentions come into the picture, they convince the agent that he really has these clear-cut, precise and explicit intentions governing its actions. In order to solve the communication problem the agent therefore creates a user-interface for itself. Dennett's take on the creation of the self and the nature of personhood shall not be discussed here due to lack of space and the necessity to treat it separately due

strong relation to this observation is Dennett's take on social interaction as is necessary in e.g. keeping a secret. To keep a secret is an important and complex intentional stance, since one has to know that p, that somebody does not know that p, that one does not want him to know that p and that one knows how to hinder him from getting to know that p. As in the case of growing complexity in the environment and its being mirrored in the complex systematic makeup of the system, social complexity can be looked at as an extension of environmental complexity from a unifying perspective. To deal with the affordances of this extension into the social realm a new kind of organization of the mind is necessary, an organization so different that it legitimizes calling it a different kind of mind. It is the possession of *language* and the ability to deal with symbols that enters the picture (Dennett 1991: Ch. 7-9, 1994).

Dennett therefore observes that as long as natural psychologists, i.e. social animals do not talk and communicate about intentions, compare notes, discuss and ask for reasons, there seems to be no selective pressure on explicitly representing those reasons (Dennett 1996b: 131). To anchor a free floating rationale to the agent to make it its own reason it must do something- it is an achievement (Dennett 1996b: 133). What has to be added to a computational architecture embodying information for it to be able to appreciate, understand and exploit the wisdom interwoven in its neural nets is *symbols*. They are movable, can be manipulated and composed into larger structures. Crucially, Dennett equates the possession of symbols with the possession of language.

Putting forward such a hypothesis, Dennett takes a high burden upon his shoulders. He has to explain what role language explicitly plays, i.e. explain how this *objectification of content via words* can be explained, i.e. how mental content can become itself object of explicit mental representation. In doing so, he has to make clear how the cultural and biological parts of this skill fit together.

To put Dennett's take on language into focus it is necessary to further explain his overall view of cognition. As Dennett states, "[e]very agent faces the task of making the best use of its environment (Dennett 1996b: 134)." As has been argued, Gregorian creatures do take in designed portions of the environment to improve their efficacy, but the primary source of human intelligence according to Dennett is the habit of off-loading as much as possible of cognitive tasks into the environment, extruding our minds into the world where we process, store and re-represent meanings, thereby freeing us from the limitations of our animal

to high dependence on other strands in Dennett's philosophy which have only been touched upon in this essay, i.e. his idea of consciousness as a virtual machine (Dennett 1991), the self as originating from a stream of narratives and Dennett's idea of the implications of anti-Cartesianism on a sound philosophy of mind (Dennett: Chapter 1-5). Chapter 8 of this essay alludes to some of these topics.

brains²⁹. A natural agent does possess biological skills and in order to solve problems he develops new skills or changes his environment, or both. One special example of a skill that Dennett notices is *labelling*, i.e. changing the environment to enhance and simplify search procedures. The simple, deliberate use of marks is hereby the precursor of writing, a way to put our designed innate tracking and pattern-recognizing talents to optimal use by spreading the mind into the world, i.e. off-loading memory and enabling the existence of a skill requiring e.g. organization and efficient search procedures in the first place. The manipulability of any system of marks, pointers, labels or symbols therefore depends on the underlying robustness of native talents in tracking and re-identifying. Most importantly, native resource management techniques do not make a difference between *internal* and *external environment*, i.e. an internal pointer or label in our mind has to fulfil similar conditions as an external pointer: if it does not suit our underlying talents well, it is worthless. Dennett's hypothesis hereby is that in a way *words* are the explanatory primitive elements of mental labelling and hence of human intelligence.

Dennett's answer as to what role language plays in human condition is furthermore what Dennett calls speculative and empirical. Words, as external pointers, are primarily to be thought of as marks or labels. Children who hear words rehearse these and build associations around them. The child thereby gets the ability to label without understanding by simply rehearsing words as labels that it has heard in some context. These stupid, unthinking practices turn into the habit of representing one's own states and activities to oneself in the following way. Anchors of familiarity of words, gained by rehearsal, give words an independent identity in the system. To serve as a useful, manipulable label in the refinement of the resources on the brain, the word must be a ready enhancer, a short cut for sought-for associations that are already in the system. The child can then turn to deliberate labelling to enhance its thought (Dennett 1996b: 149-152)³⁰.

In sum, by creating labels and attaching them to experiences we create *a new class of objects* for pattern-recognition, association-building, etc. In thinking we do not have to think all our representations over and over, but can just rely on words which themselves can then become the objects of further thinking. Language as originating in the practice of talking to

²⁹ Of course at once the question arises how the two concepts of intelligence fit together- on the one hand extruding one's mind into the environment and on the other taking in designed portions of the environment into the mind. Dennett's answer to this question is complex and will be critically discussed in Chapter 8.2.2.

³⁰ At this point in the discussion of Dennett's views the reader may be puzzled how Dennett can use any sense of 'deliberate' in his explanations and at the same time stress that original intentionality is not what is to do the explanatory job. Hereby it is helpful to notice that first the term 'deliberate' can be understood in a teleological way, i.e. the child does do something that starts to fulfil a function for it, that is aimed at something, and second, that Dennett's reconstruction of the property of being 'deliberate' or 'conscious' in itself is a complex system of thought that is supposed to explain away any 'Cartesian' elements in these terms and thereby forego criticism.

oneself and in the form of labelling becomes the foundation for explicit representation. The consequences of possession of language are therefore two-fold. First, one gains cognitive power by organizing one's own mental states into a more efficient form, and second, by creating a new class of objects whose members can become the objects of manipulation and thinking themselves. We do not only think *in* words, but also *by using* words.

7.4. Wrap-up

This chapter has been the first stage of an explication of Dennett's project to provide us with an answer to the double task of explaining both the similarities and the differences across intentional creatures. To do so I have introduced Dennett's account of the genesis of intentionality in the form of a hierarchy (Chapter 7.2). I have defended the hypothesis that Dennett changes from a predictive to a teleological concept of intentionality and hence conflates the intentional with the design stance, giving the concept of design and evolutionary explanations a necessarily important role in Dennett's philosophy.

Discussing the differences between the levels of the hierarchy I have identified the concept of representation as wanting explication. Clearing up conceptual confusion with regard to the concept of representation (Chapter 7.3.1 and 7.3.2) I have identified explicit representation to be the major point of interest in Dennett's aim to explain the singular nature of the human mind *via* the concept of language.

Giving language a primary role in explaining human intelligence evokes great expectations to the power of the concept. During the discussion therefore several questions have emerged which deserve treatment.

First, the *genesis of language* is unclear. Dennett only provides us with an account of how a child can incorporate *already existing* language, but not with an account of the development of language in the first place.

Second, in close connection to the first question Dennett has to explain how deliberate use of representation can come into play without making use of deliberate use, since the turn from unthinking to deliberate labelling as a developmental achievement by the child cannot be explained by its prior abilities of deliberate thinking without presupposing what one wants to explain.

Third, Dennett has to tell a story of how objectification of content can be explained by the concept of language as construed in his evolutionary framework. As by objectification Dennett tries to explain some of the human mind's crucial cognitive capacities, e.g. to build tools from tools or to make tools themselves the objects of thought by (re-)representation, a

lot of the plausibility of Dennett's whole account of human intentionality hinges on the plausibility of Dennett's account of objectification.

Giving answers to these questions is therefore the task of the following chapter which forms the second part of the explication of Dennett's newest concept of intentionality. The concepts of tools and memes are the instruments with which Dennett tries to give these answers. For this reason I propose the following structure for the last part of the thesis. I will explicate Dennett's *concept of tools* in general. By *doing so* I will make clear the crucial connection of the concept of tools to the *concept of memes* (Chapter 8.2). Subsequently I will critically evaluate in how far Dennett is able to answer the questions posed above (Chapter 8.2.1-8.2.2). Last, I will evaluate the relation between the concept of tools and the mind in Dennett's philosophy in general (Chapter 8.3).

8. Tools, Memes and Mind-tools

8.1. Introduction

Dennett's idea to formulate his views in terms of a philosophy of tools is rather new. But though it is new, it is not isolated from the rest of Dennett's work. The interwoven character of Dennett's whole philosophy is what makes an interpretation or evaluation of his position difficult primarily. Many of Dennett's assertions and hypotheses, if considered in isolation, seem mystic and non-understandable, and even if this impression does not always vanish when one takes the whole context of his assertions into consideration, often more sense, if any, can be made of Dennett if he is interpreted globally and holistically. Globally and holistically is meant here to refer to two considerations: first, Dennett's philosophy is an evolving system where one hypothesis depends on the other, and second, if one tends to view Dennett's ambiguous claims in the light of his main aims, topics and convictions, a lot of the ambiguity in his statements can be reduced. This is the first line I propose to take towards Dennett's philosophy of tools.

Nevertheless, an approach of such kind deserves special attention and caution, since too fast one's own views can be read into Dennett's vague style. The second line that I propose to take is to look at the best treatment of Dennett's theory of tools by Andy Clark (Clark 2002 in Clapin 2002) and show how Clark gets a lot wrong. This contrastive methodology is supposed to show the general character of Dennett's philosophy as well as to provide a foil against which a different, more cautious interpretation of Dennett may be accomplished.

As the argumentative structure of this chapter has already been explained in chapter 7.4, I continue the explication without further ado.

8.2. Tools, Mind Tools and Memes

What, then, are tools according to Dennett? Dennett defines tools to be designed portions of the environment that are taken in by Gregorian creatures which are thereby endowed with intelligence that is embodied in the tool (Chapter 7.2). Tools help to instantiate intelligent behaviour and solve problems by adjusting them to our biological skills and by providing the basis to build further tools. As to what exactly constitutes the set of tools Dennett's remarks remain rather vague.

A tool can be whatever makes, to gloss it sloppily, life easier to the user. A can opener is a tool, since it makes it easier to open cans otherwise than with one's hands. A rake is a tool since it helps the user to change his environment, in this case the soil, in such a way as to allow for easier living, e.g. by growing crops (Clapin 2002: 95-96). Shopping lists and calculators are also tools, since they make the problems of remembering and calculating easier. Furthermore, the set of tools also encompasses entities like slow-motion cameras, since these fit problems that are presented to humans in a speed which fits our natural tracking abilities better (Dennett 1996b: 138-144).

The most important tool from the perspective of explaining intelligence is language (Chapter 7.3.3). But according to Dennett, language is not only a *tool*, but also a *mind tool*. In general, Dennett does nowhere give an explanation of the difference between the concepts. Can therefore the two notions be used interchangeably? Dennett's commentators criticised that if this were the case, as a presumably disastrous result Dennett's use of the notions 'mind' and 'agent' also would be interchangeable.

Tools like hammers or can openers are 'out there in the world'. They are *separate* from us, they do not *constitute* us. We *use* tools, but we *are not* tools. If somebody asked one what it were that he was, he would not like to answer that he is his hammer or his can opener. The concept of an agent does not include external items that we, as agents, *use*.

In spite of this, tools on Dennett's account *do* constitute the agent. Our mind is supposed to be made of the use of e.g. mental shopping lists, words and tricks to remember better. Dennett's conceptualization of tools is therefore at least unintuitive. But even worse, it is even judged to be empty. Extension of cognition to whatever one likes to call a tool creates the worry of losing the notion of an agent in a welter of ubiquitous external cognition. If cognition does also encompass calculators, rakes and alphabets, the agent's mind and thereby

the agent himself can be nearly everything and every time one uses a different tool somebody else. Hence Dennett is found in need of providing for a difference between tools and mind tools, where *mind tools* are to constitute the agent, but *tools* not.

This way to criticise Dennett seems to be highly warranted on first sight. Dennett remains unimpressed by it. As Dennett formulates it, the worry foots on a Cartesian view of the mind, i.e. unwarranted presupposition that there is something like a thinking centre, a nugget in every individual, the thinking mind that constitutes the origin of thought, intentionality and consciousness. Dennett believes that the force of this presupposition mainly comes from its ubiquity, not its correctness. Therefore Dennett is a critic of what he calls one of the main tenets of classical cognitive science, namely the way in which content is isolated and individuated in a system (Dennett forthcoming). According to Dennett, the primary unit of analysis in cognitive science is some kind of internal nugget, a central control system connected to the environment by transducers and effectors. He offers an alternative account. His primary unit of analysis is not a central nugget, but the whole system including bodily form and local context. To Dennett cognition is ubiquitous. This view is in accordance with his idea of a system possessing knowledge or being in states describable as content bearing states due to their carrying information (Dennett 1982c, Chapter 7.3.2), not only when possessing explicit representation. Content can be, as has been argued by Dennett while constituting his teleological theory of content, attributed whenever it is apt to describe an entity as having a function, i.e. as being good for something. Thereby the notion of cognition is altered to encompass all the contextual and systemic features a creature possesses that promote adaptive success, including all tools.

From this point of view, the extension of cognition is not a problem for agency. Whatever possible difference may be construed between tools and mind-tools, it is *not* to be found in differences in content, adaptive success or aptitude to description in representational idiom.

The loss of the agent isn't a problem for Dennett from the very beginning. Since to Dennett the agent and his consciousness are just a virtual machine that tells itself a story (Dennett 1991: Ch. 9), may it be as a side effect of the development of communication (Chapter 7.3.3) with others or with oneself so as to make locally present information in segmented mind modules available to other modules (Dennett 1991: Ch. 7.5), Dennett does not care to loose the agent. According to Dennett it is an illusion anyways, and an illusion

responsible for philosophical confusion above that³¹. Since he does not presuppose the notion of an agent, he feels free to reinvent the agent in his framework out of this origin-less welter by the use of heterophenomenological data, the idea of consciousness as a virtual machine and most importantly the user illusion (Dennett 1995).

I therefore view Dennett's usage of the concepts of tools and mind tools as interchangeable in any of the ways that Dennett believes to be relevant to his philosophy. I propose to follow Dennett in this interchangeable usage.

This criticism concerning the extension of cognition has by far not been the only one. I would like to draw attention to a further problem regarding the individuation of tools and which I would like to use to change to Dennett's evolutionary perspective on tools that will set the stage to enable me to answer the questions concerning Dennett's account left open in the last chapter (Chapter 7.4).

A tool is defined as a designed portion of the environment a user can use to increase his adaptive success. Hence the concept of tools is closely bound to the concept of adaptive success. The philosophical concept of adaptive success itself is rather broad, encompassing several contradictory ideas as of *to what* adaptive success can be applied. This creates problems for Dennett's definition of a tool in terms of adaptive success.

As an example, imagine a user spending his whole day calculating random sums with his calculator. On the one hand this user can naively be described as using a tool, i.e. the calculator, but on the other hand his doings cannot be described as increasing his adaptive success in any meaningful way. It seems that either the user uses a tool that does *not* increase his adaptive success, or the situation isn't apt to be described as tool use in spite of its intuitiveness.

What is Dennett's answer to this problem? It might seem that Dennett would want to provide a standard against which adaptive success would have to be measured in order to define tools, e.g. an entity is a tool if it in most cases/ often/ usually increases the user's fitness. Such a position, though, would be highly unstable, since the concepts of 'most often' and 'usually' are hard to be given a clear definition and would thereby leave many cases in a grey shaded zone between being a tool and being counterproductive, non-intelligent activity.

Dennett's way to evade such an unstable position is neither to ascribe adaptive success to populations of users in the sense of normally or usually successful, nor to the individual

³¹ Furthermore, questions as to moral responsibility are reconstructed by Dennett pragmatically, but a full discussion of these implications cannot be given in this context due to lack of space. The interested reader may want to consult Dennett 1984.

itself in the sense of favouring his survival, but *to the designed portions of the environment themselves*, the tools and skills that we learn in social and cultural context. Tools themselves are viewed as the objects of evolution and Dennett can evade criticism regarding the concept of adaptive success as applied to populations or individuals by introducing a new object of evolutionary selection: the well known concept of a *meme*. Dennett's meme theory (Dennett 1995, 1998, 2000a, 2000b, 2002) finds its roots in Richard Dawkins' (Dawkins 1976, 1982) theory of memes, defining memes as bits of cultural information that reproduce themselves by using human brains as hosts³².

Tools as designed portions of the environment can thereby be viewed as a class of memes, i.e. the class of memes that confers to the host adaptive success. Such a move does reconstitute the individual as the object to which the concept of adaptive success can be applied, but at the same time allows and explains cases in which the use of tools is counterproductive to the survival of the individual. In the latter case evolutionary success is better ascribed to the meme, the ability to use a calculator *itself*, rather than to the user. Of course, as memes are virus-like entities and dependent in their own survival upon the survival of their hosts, adaptive success of both memes and creatures invaded by these memes will often co-occur, but *only* due to the indirect dependency of the fitness of the host as increased or lowered by the memes it carries.

Neither Dennett's theory of tools nor his theory of memes is very original, but the way in which Dennett puts these ideas to work is unique. Using the concept of a tool Dennett is able to extend the notion of human cognition to the external, designed portions of the world. Using the concept of a meme, Dennett is not only able to provide an account immune against usual criticism regarding adaptationistic individuation of tools, but more importantly is able to give an answer to the open questions from the last chapter. I would like to evaluate critically Dennett's answers to these questions with the background of his tool concept in the following.

8.2.1. Genesis of Language and Deliberate Use

The main limitation on answering the first question, i.e. the origin of designed portions of the environment, e.g. words, is that it has to be explained without presupposing prior intelligence to invent them. Giving an answer of this kind amounts to evading what Clark calls 'the paradox of active stupidity' (Clark 2001: Ch. 8).

³² In the following discussion I propose to use the definition of a meme as a bit of cultural information that reproduces itself by using human brains as hosts in spite of its broad sense, since for the applicability of the arguments given in this paper a more exact definition is not necessary. More formal and narrow definitions bring in conceptual problems in addition to the problems presented here and cannot be discussed due to lack of space. A discussion of more formal and narrow definitions of a meme including the respective conceptual difficulties can be found in Rose's 'Controversies in Meme Theory' (Rose 1998).

“This would be the idea that making the moves that sculpt the environment so as to allow *cheap* problem-solving itself requires *expensive*, advanced, design-oriented cogitation. The threat being that only *clever* brains could make their world smart so as that they could be dumb in peace- a result that would deprive the tool-based scenario of its appealing role in both originating and partially constituting advanced, reflective thought and reason (Clark 2002 in Clapin 2002: 70).”

Dennett’s concept of the origin of tools observes this limitation by introducing the notion of ‘found objects’ and by meme theory as *two perspectives* on the same phenomenon.

The concept of ‘found objects’ takes the *perspective of the user*. It supposes tools to be found by chance. A user ‘finds’ a tool by chance without looking for it, e.g. by chance he starts to hit one fire-stone on another and finds that by doing so he can light a fire. Being stupid, he does not search for a way to light a fire, but having found it he has become smarter. Tools in general are envisaged to be found in this way. Since one human being can imitate another and learn from him, the tool is transmitted from one being to another and so becomes common. Dennett’s concept of the tower of generate-and-test which provides the basis for his idea of the development of minds foost on this idea and its extension to the concept of tools is therefore only natural.³³

Taking the *perspective of the tool*, tools themselves are objects of natural selection via differential reproductive success. The designed and adaptive character of objects in the world in relation to the user hinges on the reproductive success of the *meme*, thereby freeing the user from the necessity of deliberate design by deriving smart design indirectly *via reproductive success of memes*. The user himself does not have to be smart, since natural selection between memes does the selection of smart tools for him.

In the case of pure biological evolution the tool use develops as an outcome of adaptive success of the individual, but in the case of cultural evolution the tools develop from adaptive success of the *meme*. Hence which tools are used, how one’s mind is organized and which actions are chosen is *partly* up to the tools and *partly* up to the user in the sense that external portions of the environment can be taken in to furnish one’s inner environment in multiple ways. Differential reproduction of memes in the sense of slightly differing versions invading human minds thereby allows for much faster evolutionary development than through pure biological selection. Whereas in biological selection the success of a property being possessed by an individual needs at least the time of one possible biological reproduction to

³³ This extension is exemplified by the workings of the mind of Popperian creatures (Chapter 7.2). In the minds of Popperian creatures possible options for action are evaluated, i.e. possible actions are generated, tested in an internal environment, and chosen.

show itself, in the case of cultural evolution selection can take place rapidly and continually in one's head.

In a similar fashion Dennett also answers the second question, i.e. how deliberate use of representation can come into play without making use of deliberate use, since the turn from unthinking to deliberate labelling as a developmental achievement by the child cannot be explained by its prior abilities of deliberate thinking without presupposing what one wants to explain.

According to Dennett, labelling as a *meme* uses the child for its own instantiation and reproduction and, by being provided by competent speakers, infects the child with itself. The ability of language to do so foots in the child's biological abilities, but the child's doing so is dependent on the existence of language as a meme. The child's development of intelligence is therefore partly due to a biological basis, and partly, but also crucially, due to the existence of culturally transmitted and adapted skill or information packages in the form of the social practice of language. To sum up, 'deliberate' in the sense of 'being aimed at something' as a property of thought is thereby reconstructed as emerging from the cognitive machinery that is made up of memes installed in our minds. In metaphorical words one can say that we act deliberately because the memes do the thinking for us. 'Deliberate' in as a property of thought in the sense of 'conscious' according to Dennett emerges basically in the same way, but this topic needs separate treatment and will not be discussed in this context.

8.2.2. Objectification of Content

Dennett's account on the origin of language and the turn from non-deliberate to deliberate labelling is prima facia convincing. Unfortunately this is not already the whole story that Dennett has to tell. He stills owns us an answer to the third question, i.e. how objectification of content can be explained. Dennett's answer to this question is more complicated and complex, but worth discussion since Dennett's whole concept of human intelligence hinges on providing a convincing account of the objectifying power of language.

I want to claim that Dennett's answer regarding the question of the objectifying nature of language is not convincing. To prove my hypothesis I will first explain Dennett's concept of purportedly anti-Cartesian content fixation in the tool approach and then show that Dennett's account of language runs danger to injure Dennett's limitations of doing anti-Cartesian philosophy or remains unclear.

The mind is a collection of tools that are supposed to be on par, that is, there is supposed to be no special, single mind code like Mentalese, no procedural 'nugget' analogous

to a CPU of a computer, no real cognitive machine to which the rest of cognition adheres peripherally. Furthermore, tools are supposed not to have a user, but to essentially and exclusively constitute the mind through their working, intentionally undermining the division between ‘thinking in’ and ‘thinking with’. Though tools may be the users of other tools, it does not follow that there exists any tool that uses all other tools in a way that would determine the intentional status of the tools the ‘supra-tool’ would use.

Since on Dennett’s account content and aboutness get in the picture through skilled engagement between the agent and the world, no single, basic or fundamental mind tool can support all the observable phenomena and inner processes that have to occur in order to make talk about understanding or grasp of meaning appropriate. Since this kind of holistic understanding depends on the activity of multiple, non-privileged tools, “there is no more reason to treat internal representations of language as ‘merely derivatively contentful’ than to treat a symbol in [M]entalese or any other mind-tool”(Clark 2002: 66-90). No single tool has intrinsic, original aboutness, but only the holistic whole of thought does.

To explain objective use of content Dennett therefore endorses some kind of ‘bootstrapping externalization story’ that puts language as the main ‘objectificator’ into place. *Somehow* this externalization is supposed to work, and *somehow* it is supposed to “augment or transform” human thinking so that in the end thinking about thinking and consciousness emerge. The problems lie in the notions ‘*somehow*’ and in the considerable difference between “*to augment*” and “*to transform*”. Assuming tacit mechanisms to be enabling externalization, the question arises, what the innate, biological mechanisms are in the first place, and how the relation between internalization and externalization can be thought of from this perspective.

This point may be made clearer if one looks at Clark’s revision of Dehaene’s (Dehaene 1997, Dehaene et al. 1999) work on human arithmetical skills as a model for a model of language. Dehaene supposes numerical cognition to depend on three factors. First, an innate, biological sense for low grade approximate arithmetic, second the culturally acquired capacity to think about exact quantities in courtesy of verbal and language specific representations of numbers, and third a cultural evolutionary process involving the use of body parts as stand-ins for numbers. The genesis of exact arithmetic is supposed to start with the innate fuzzy arithmetical sense and the use of body parts for numbers. Then, accidentally (to avoid the paradox of active stupidity) the agent discovers that three plus one finger are four fingers and then through a *miraculous* process develops a ‘virtual machine’ that can use linguistic tags, i.e. number words, to think mathematically. Arithmetical skills therefore

depend on the discrete, digital use of an analogue counting device built into our brains. How satisfactory is this example?

On first sight it does a good job in supporting Dennett's thesis that for advanced thinking through internalization no simple translation from a public to a biological code is required, so that the unacceptable Mentalese picture à la Fodor (Chapter 6.2.1, 6.2.2) is skilfully avoided. But also crucially the presence of *links and bridges* between externalization and internalization is stressed as being essential and necessary. Then, one may again ask: how do these links and bridges look like? To answer that this depends on the internalization of externalized contents would create nothing except a vicious circle. To avoid this fallacy, it is necessary to put forward an account of how the human mind can use *externalization*, and especially how it can use internalization to get the benefits from externalizations back in. This account has to explain how the initial internal biological mechanism functions without presupposing it to be able to use externalization by fiat. Dennett's idea hereby is to break down complex tasks of externalization and internalization into tasks that can be fulfilled by biological pattern matching and recognition mechanisms making use of found memetic tools.

Moreover, the account has to explain how this specific mechanism is able to use a specific kind of externalization to gain a specific success, and I want to argue that there is a decisive difference between the general principle of externalization and externalization as it can be used by a specific tool. As language is our main topic here, I propose to stick to language and show how the general observations just made apply here.

The problem with the usual boot-strapping accounts of objectification and representation generation is that they suppose language to be the objectifying mechanisms, but such an argument can be viewed as both too strong and too weak. It is too strong, since it uses a general device such as language to explain a specific ability. Of course, taking such a powerful tool as language with its obviously huge range of applicability for granted, makes it easy to explain a simple task a fortiori. Problematically, it seems highly improbable that such a mind tool might have existed in the first place, and Dennett denies such a stipulation, explaining language itself as a gradually and incrementally working mind tool that co-evolved with other mind tools. I do not want to open the vast range of problems and speculations about the genesis of language and externalization at this point, but draw attention to how the argument above may be too weak, requiring a short digress.

Language is as Dennett observes highly systematic, combinatorial and productive. It is a mind tool of immense power in the sense of range of applicability and generality, in so far as it does not stick to a specific pragmatic or semantic domain. I can talk about buying eggs

and cheese, and I can talk about God, the immaterial and my coffee. *How could the status of such a mind tool in the bags of mind tools be looked at?* Let us assume that since language had to develop somehow, it must have developed from a simple to a complex form. The complex form is the one today, the simple one the topic of further analysis.

This simple form must have developed, accidentally, as a found object through trial and error. It has to be observed, however, that this language, however simple or complex it might have been in the first place, was surely very narrow in its applicability. It seems at least highly speculative to assume a device like language with broad applicability to pop out from evolutionary history. But, then the development into a more complex form as it comes to applicability has to be looked on as either developing the one language mind tool using other tools further and further, or putting several mind tools together under some kind of usage by language.

If the examples given above for what I can talk about may be too far fetched, I could add that I am also able to talk about my visual experience, my bodily position, my plans for the future and my past history to give them more of a internal nature, not overtly presupposing culture or other external systems or objects that are supposed to somehow work in the ‘externalization business’³⁴. The latter possibility has to assume that in the development of the one mind tool responsible for language, the range of applicability must have increased, and as the examples show, it must have put to work a lot of other mind tools that the mind comprises, e.g. the ones having content about visual states or the ones holding content about bodily posture. Such a general device in the end would have to be a “universal translator” that is able to interpret all the different non-propositional mind tools on par in order to make sense of them all. This might or might not include a translation into a generic code, and even if it didn’t in the sense of Mentalese, it seems at least hard to me to imagine how such a complex device could look like. Nevertheless, one could put the point in a nutshell by saying that instead of presuming a universal mind-code, Dennett gets rid of the code and instead assumes a universal interpreter.

The fact that this does not solve, but move the problem, is obviously clear. Now, Dennett does not have to assume this and can put the second possibility into work, namely that several mind tools on par, together, constitute the objectifying nature of language. This might be always the same mind tools or changing ones, according to the needs of the organisms.

³⁴ This change of examples is supposed to leave out cultural and what could be called externally contextual elements as far as possible. I stipulate that many would intuitively agree that feelings, visual and proprioceptive experience are in a sense epistemologically simpler and nearer to human cognition than e.g. the Immaterial.

On first sight, such a move gets rid of a universal supra structure, a ‘supra nugget’ that is controlling language as appeared to be necessary to make sense of the first possibility, and as Dennett looks on language as essential for consciousness and human culture. But then, observing the already noticed generality, combinatorial nature and productivity of language, but here especially its coherence³⁵ - talking about different things we for sure do use different words, but intuitively do not change the language- would have to emerge as a result of the structure of the underlying mind tools, not as their controlling mechanism. Such a structure would have again to be highly complex, and I do not see many possible arguments to look at the single mind tools contributing to the overall emerging language as some kind of code. The items of this code may be individuated by single skills, but such a view would be highly redundant, because a skill does not deal with just one specific situation, but with unaccountably many situations in a specific domain, that is tasks have to be thought as type devices to make any sense in an ill-defined and dynamic environment³⁶. This point alone does for sure not license the assumption of necessity of a mental code, and an interpretation through dynamic system theory may suit the purpose way better³⁷, but in the end of the process and interaction of all the contributing mind tools something that is isomorphic to natural language would have to emerge- remember, we cannot assume a translator or interpreter in this account without falling back into possibility one. How this isomorphism may look like and how it possibly could work out to create such a huge device as language I cannot conceive. The worry that some kind of universal principle would have to be at work does not seem easily eliminable.

To put these consideration back into the main context in answering the question how the ‘boot-strapping story’ about the objectification function of language might look like, I would like to say a word or two about the ‘quantity’ of this analysis and summarize.

First, taking language as the primary device doing the objectification job raises the question how this could be done without assuming a universal interpreter or some other kind of mystical universal principle holding the tools together. One may answer that the point is not to look at language as the primary explanatory tenet, but on the underlying tacit, non-

³⁵ By coherence I mean the property of language of staying the same when changing the semantic domain. Taking about molecular physics and talking about a sunny afternoon for sure amount to the use of *different words*, but in an intuitive manner we can still say that both of the accounts can be delivered in the *same language*. This is not only meant to be a point about grammatical structure, but also in the sense that one can talk about a sunny afternoon in terms of molecular physics and vice versa, qualitative worries being granted.

³⁶ This problem can be exemplified using colours: I can talk about green and about red and all other colours existing in our phenomenology, but assuming a tool for each colour would make the number of necessary tools explode.

³⁷ Dynamic system theory in the domain of mental representation can be put to work in a Clarkian manner (Clark 1997), but the discussion of this possibility is beyond the scope of this essay.

propositional mechanism. This answer, though, falls short on the fact that Dennett does use language in order to explain objectification. Even if one ignores this point, it then would be necessary to show what other kind of mechanism could do the job, may it be tacitly, using propositional content or non-propositional content, using language or not. The jump from tacit, domain specific skills to a full blown language and use of items worth calling propositional mental representation is simply too huge to be explained by ‘*somehow*’. My analysis does not show that Dennett’s account is false, or that a universal inner code is mandatory for objectification, but shows where explanation is missing. Granting the obvious existence of human use of externalization, and even granting culture and language its importance in an explanation of human behaviour, it still is an *open and important question*, how these are supposed to function in the mind-world relationship³⁸.

8.3.A ‘Kind of Mind’

As Chapter 8.2 has made clear, Dennett’s concept of tools is conceptually shaky, i.e. it leaves open crucial questions regarding the role tools play in the explanation of the human mind. Since Dennett’s philosophy of tools is new, empirical and hypothetical, such a finding is not very surprising. Dennett’s theory has not have had time to mature yet. But it is problematic, especially when one tries to get a clear picture of what exactly Dennett’s philosophy amounts to. This is both true regarding technical questions as discussed above, but also regarding more general questions. Commentators of Dennett like Clark have therefore often put forward hypotheses regarding Dennett’s philosophy that are shaky themselves. Since Clark’s discussion of Dennett’s concept of the mind as it comes to tools is probably the most exhaustive and influential one, I would like to show how Clark goes wrong in re-conceptualizing Dennett. This argumentation will lead me to a final statement about the success of Dennett’s philosophy of tools in explaining the uniqueness and special stand of the human mind in Dennett’s hierarchical model of intelligence.

The biggest danger to Dennett is to fail to provide an alternative account of intentionality and end up somewhere in Cartesian realm, e.g. by nevertheless having to provide for a ‘central nugget’ of content fixation. How Dennett tries to evade such an ending in detail has

³⁸ It is interesting to notice how the location of Dennett’s own position has changed with respect to the RTM. Whereas in earlier writings Dennett stressed the point that though some kind of RTM may be correct, this by far is not *obvious*, Dennett today can be interpreted as holding a position which can itself be said to allow for some kind of RTM, though it does not do so in obvious ways but in a way that needs *further* explication. It is therefore possible to say that Dennett’s theory did not bring the field much further, at least when it comes to making some kind of RTM plausible.

remained rather vague and I have tried to show that at least his project has not been fully successful (Chapter 8.2.2). Hence Clark supposes that if Dennett wants to get rid of Cartesianism, he has to do so radically and completely. As Dennett introduces tools to do this job, Clark stipulates that Dennett has to make the mind to be *completely* composed of tools- *it is tools all the way down* (Clark 2002 in Clapin 2002). There shall not be anything left except tools in the mind, nothing that could create doubts regarding its anti-Cartesian standing. Dennett himself invites such an interpretation by glossing the human mind to be ‘a bag of tools (Dennett, 1996*b*), or a ‘bag of tricks’ (Dennett 1991: 280). Though Dennett seems to invite such interpretation, I stipulate that it is either wrong or only superficially right.

To say that our human minds are *exclusively* made of tools on first sight seems obviously wrong. For sure it is wrong when tools are supposed to play a role only in representational schemes, as Clark supposes (Clark 2002: Ch 8), since it is one of Dennett’s main point that human minds are *not* just representational schemes. But even if the concept of tools is supposed to encompass rakes and can openers, it seems clear that we are not made up of rakes and can openers, since a human being without all these tools can still be imagined to be a human being with a mind. Furthermore, if one follows Clark in believing that tools are a subclass of memes, minds of course also cannot completely be composed of tools, since we are not completely composed of memes³⁹. In addition, the use of tools presupposes biological skills which themselves at least were not termed tools by Dennett.

On all these interpretations the claim that human minds are composed of tools is blatantly wrong, so that charity obliges one to take a different interpretation of the claim. On such an interpretation to say that we are tools all the way down means that what we are, as conscious, cultural, social human beings in the sense of what makes us different from animals, special due to the possession of tools, and that it is only tools that make us so special. This line of interpretation seems also better when judged against Dennett’s broad philosophical project and perspective.

Dennett stresses the concept of tools, since tools fit in nicely with a skill based account of intentionality. Furthermore, to say that our special human minds are just tools is also an expression of Dennett’s anti-Cartesianism that trades in intrinsic intentionality, aboutness and a central content fixation nugget for a user-less bag of tools and a virtual machine. One just shouldn’t overdose Cartesianism. Moreover, if our minds are supposed to be special for the kind of consciousness we enjoy, the idea of tools constituting the machinery behind this consciousness also connects with Dennett’s idea of consciousness as a user illusion and a

³⁹ One may also note that the concept of a meme even *presupposes* the concept of a host and therefore an explication of the host being purely composed of memes to be begging the question.

virtual machine installed in human brains as a result of the possession of language, which, of course, itself is a tool. In addition, to view our kinds of minds as being composed of tools also aligns with Dennett's point on the externalization of the mind both in the sense of external cognition as well as in the sense of our minds being invaded by memes in all their different forms, i.e. social practices, laws, ways of thinking, etc.

If one glosses Dennett therefore as stating that our minds are tools all the way down, it means to draw attention to the fact that Dennett tries to show what we said was a burden on him, i.e. what makes human minds so *special* as to legitimize their belonging to a unique kind of mind. But we are not *just* special, i.e. not just tools, but we are obviously more.

8.4. Wrap-up

With this chapter I have finished the second stage of an explication of Dennett's project to provide us with an answer to the double task of explaining both the similarities and the differences across intentional creatures. To do so I have introduced the Dennettian concept of tools (Chapter 8.2) and shown its crucial connection to the concept of memes. Particularly, I have shown how Dennett combines the concept of memes and tools in order to create a basic framework for the explanation of the genesis of language (Chapter 8.2.1). Using the same concepts I have elucidated Dennett's view on the turn from non-deliberate to deliberate use of tools (8.2.1). Though Dennett's *general* view on these issues seems convincing, in Chapter 8.2.2 I have shown that for the crucial concept of language as a tool Dennett has not been able to provide a convincing explanation.

Has therefore Dennett's project to provide us with an account of the special status of intentionality failed? In a way it has, since many questions have remained open. On the other hand Dennett's concept of tools is so new, empirical and hypothetical that drawbacks and developments are to be expected. Besides, Dennett himself would not be very surprised by his partial failure, and not even very worried. To him explanations of human intentionality are a field that cannot be answered in any other way than through macro-reductive thinking in some way or other, i.e. by *exploration of new ideas*. In which way *exactly*, history will show. According to Dennett, there is no alternative to trying to flesh out new ideas, scientific findings and novel concept if one wants to explain human intentionality at all. We simply do not know enough about the human mind to put up a theory of human intentionality yet. Dennett has already pushed the field in this direction, and as I have tried to show, there is a lot left to be pushed further.

9. A Few Final Words

Looking at the development of Dennett's theory several perspectives are possible. One is to state that Dennett's early philosophy has nothing to do whatsoever with the later one. Another is to stress Dennett's more general aim of naturalizing the mind and to view the different approaches Dennett presents as different instantiations of this endeavour. In this thesis I have tried to do both and none and to get the best out of it- on the one hand stress the diachronic coherence of Dennett's position in naturalizing the mind in order to show the mutual connections and motivations of the positions Dennett took over time, and on the other view Dennett's development from the early to the late philosophy first as an evolving answer to criticism and error correction, and second as a development that includes major changes of emphasis, perspective and interest.

In so far as Dennett is a philosopher informed by the natural sciences, his own philosophy follows the newest finding both in their successes and their failings. I have shown that Dennett's theory offers a lot of explanations to the questions that we have about the nature of the mind, and especially the human mind in a unique and fascinating way of doing philosophy. In many aspects it is different from the canonical form of analytic philosophy and that's what makes it so fascinating and exciting. Dennett is not a philosopher that is standing still- he is moving and trying to push the field forward. On the other hand I have argued that Dennett's philosophy, and especially his newer writings have to be considered critically and with care. I have identified and discussed a multitude of problems in Dennett's account with respect to the coherence of his conceptualization, his perspective on language and tools, external cognition and representation. Nonetheless, in spite of all criticism I hope to have been able both to convey to the reader the sense of fascination that Dennett can excite.

10. References

- Akins, K. (1996). 'Lost the plot? Reconstructing Dennett's Multiple Drafts Theory of Consciousness'. *Mind and Language*, Vol. 11(1): 1-43.
- Baker, L.R. (1989). 'Instrumental Intentionality'. *Philosophy of Science*, 56(2): 303-316.
- Block, N. (1983). 'Psychologism and Behaviourism'. *Philosophical Review* XC(1): 5-43.
- Brentano, F. (1874/1973). *Psychology from an Empirical Standpoint*. Ed. O. Kraus. English edn. Ed. L. L. McAlister. Trans. A.C. Ranurello, D. B. Terrell, and L. L. McAlister. London: Routledge & Kegan Paul.
- Clapin, H. (2002). *Philosophy of Mental Representation*. Oxford: Clarendon Press.
- Chisholm, R. (1956). 'Sentences about Believing'. *Aristotelian Society Proceedings*, 56: 125-48.
- Churchland, P. M. (1981). 'Eliminative Materialism and the Propositional Attitudes'. *Journal of Philosophy*, 78: 67-90.
- Clark, A. (1997). 'The Dynamical Challenge'. *Cognitive Science*, 21(4): 461-81.
- Clark, A. (2001). *Mindware: An introduction to the Philosophy of Cognitive Science*. Oxford: Oxford University Press.
- Clark, A. (2002). 'Minds, brains and tools', in Clapin, H. (ed.), *Philosophy of Mental Representation*. Oxford: Clarendon Press, 66-90.
- Cosmides, L., Tooby, J. (1992). 'Cognitive adaptations for social exchange'. In J. H. Barkow, L. Cosmides, & J. Tooby (eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York, NY: from my point of view Oxford University Press, 163-228.
- Davidson, D. (1973) 'Radical Interpretation', *Dialectica*, 27, Repr. in D. Davidson. *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 2nd edn, 2001.
- Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.
- Dawkins, R. (1982). *The extended phenotype*. Oxford: Oxford University Press.
- Dehaene, S. (1997). *The Number Sense*. Oxford: Oxford University Press.
- Dehaene, S., Spelke, E. Pinel, P. Stanescu, R. and Twiskin, S. (1999). 'Sources of Mathematikal Thinking: Behavioral and Brain Image Evidence'. *Science*, 284: 970-4.
- Dennett, D.C. (1969). *Content and Consciousness*. London; Routledge& Kegan Paul.
- Dennett, D.C. (1971). 'Intentional Systems'. *Journal of Philosophy*, 68(4): 87-106. Repr. in *Brainstorms* (Dennett 1978b).

- Dennett, D. C. (1976). 'Why The Law of Effect Will Not Go Away'. *Journal of the Theory of Social Behavior*, 5: 169-187. Repr. in *Dennett Brainstorms* (1978b).
- Dennett, D.C. (1978a), "Artificial intelligence as philosophy and as psychology", in Ringle, M. (ed.), *Philosophical Perspectives on Artificial Intelligence*, Atlantic Highlands: NJ.
- Dennett, D.C. (1978b). *Brainstorms: Philosophical essays on Mind and Psychology*. Montgomery, VT: Bradford Books.
- Dennett, D.C. (1978c). 'How to Change your Mind'. Ch. 13 of *Brainstorms* (Dennett 1978b).
- Dennett, D.C. (1978d). 'On giving Libertarians what they want'. Ch. 15 of *Brainstorms* (Dennett 1978b).
- Dennett, D.C. (1978e). 'Reply to Arbib and Gunderson'. Ch. 2 of *Brainstorms* (Dennett 1978b).
- Dennett, D.C. (1978f). 'Skinner skinned'. Ch. 4 of *Brainstorms* (Dennett 1978b).
- Dennett, D. C. (1981a). 'Three kinds of intentional psychology'. In R. A. Healey (ed.), *Mind, Psychology, and Reductionism*. Cambridge: Cambridge University Press. Repr. in *The Intentional Stance* (Dennett 1987a).
- Dennett, D.C. (1981b). 'True believers', in A.F. Heath (ed.), *Scientific Explanation*. Oxford: Oxford University Press. Repr. in *The intentional Stance* (Dennett 1987a)
- Dennett, D.C. (1982a). 'Beyond belief'. In Woodfield, A. (ed). *Thought and object*. Clarendon Press, Oxford. Repr. in Dennett 1987a.
- Dennett, D.C. (1982b) 'Making Sense of Ourselves'. In J.I. Biro and Robert W. Shahan, (eds.), *Mind, Brain, and Function: Essays in the Philosophy of Mind*, 63-81. Norman: University of Oklahoma Press, 1982. Reprinted in Dennett 1987a.
- Dennett, D.C. (1982c). 'Styles of Mental Representation'. *Proceedings of the Aristotelian Society*, LXXXIII. Repr. in Dennett 1987a).
- Dennett, D.C. (1983). 'Intentional systems in cognitive ethology: The "Panglossian paradigm" defended. *Behavioral and Brain Sciences* 6: 343-390.
- Dennett, D.C. (1984). *Elbow Room*. Oxford: Oxford University Press
- Dennett, D.C. (1987a). *The Intentional Stance*. Cambridge, Mass.: MIT Press.
- Dennett, D.C. (1987b). 'Midterm examination: Compare and Contrast'. Ch. 10 in *The Intentional Stance* (Dennett 1987a).
- Dennett, D.C. (1987c). 'Reflections: Instrumentalism Reconsidered'. Reflections to Ch. 3 of *The Intentional Stance* (Dennett 1987a).

- Dennett, D.C. (1987*d*). 'Reflections: The Language of Thought Reconsidered'. Reflections to Ch. 6 of *The Intentional Stance* (Dennett 1987*a*).
- Dennett, D.C. (1987*e*). 'Reflections: When Frogs (and Others) Make Mistakes'. Reflections to Ch. 4 of *The Intentional Stance* (Dennett 1987*a*).
- Dennett, D.C. (1989). 'The Origins of Selves'. *Cogito*, 3: 163-73. Repr. in Kolak, D. and Martin, R. (eds.). *Self & Identity: Contemporary Philosophical Issues*. New York: Macmillan.
- Dennett, D.C. (1990). 'The Interpretation of Texts, People and Other Artifacts.' *Philosophy and Phenomenological Research, Supplement*: 177-94. Repr. in Losonsky, M. (ed.), *Language and Mind: Contemporary Readings in Philosophy and Cognitive Science*. Oxford: Blackwell.
- Dennett, D.C. (1991). *Consciousness Explained*. Boston, MA: Little, Brown & Co.
- Dennett, D.C. (1992). 'The Self as the Center of Narrative Gravity'. Kessel, F., Cole, P., Johnson, D. (eds.), *Self and Consciousness: Multiple Perspectives*. Hillsdale, NJ: Erlbaum.
- Dennett, D.C. (1993). 'Back from the Drawing Board'. In Dahlbom, B. (ed.), *Dennett and his critics: demystifying mind*. Oxford: Blackwell.
- Dennett, D.C. (1994). 'The Role of Language in Intelligence'. In Khalfa, J. (ed.), *What is Intelligence. The Darwin College Lectures*. Cambridge: Cambridge University Press.
- Dennett, D. C. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon and Schuster [DDI].
- Dennett, D.C. (1996*a*). 'Granny versus Mother Nature- No Contest'. *Mind and Language* 11(3): 263-269.
- Dennett, D. C. (1996*b*). *Kinds of Minds*. New York: Basic Books.
- Dennett, D.C. (1998). 'Memes: Myths, Misunderstandings and Misgivings'. (Draft for Chapel Hill, October 1998, available as a preprint at ase.tufts.edu/cogstud/pubpage.htm).
- Dennett, D.C., Densmore, S. (1999). 'The Virtues of Virtual Machines'. *Philosophy and Phenomenological Research*, Vol. LIX(3): 747-761.
- Dennett, D.C. (2000*a*). 'From Typo to Thinko: When Evolution Graduated to Semantic Norms'. (Draft: Memes paper for Fryssen, July 2002, available as a preprint at ase.tufts.edu/cogstud/pubpage.htm).
- Dennett, D.C. (2000*b*). 'Making tools for thinking', in D. Sperber (ed.), *Metarepresentations: A Multidisciplinary Perspective*. Oxford: Oxford University Press.
- Dennett, D.C. (2002). 'The new replicators'. In Page, M (ed.), *Encyclopedia of Evolution*. Oxford: Oxford University Press.

- Dennett, D.C. (2003). 'Who's On First? Heterophenomenology Explained'. *Journal of Consciousness Studies*, 10(9-10).
- Dennett, D.C. (forthcoming). 'Things about Things'. (Draft for Lisbon conference on Cognitive Science, May 1998), available as a preprint at ase.tufts.edu/cogstud/pubpage.htm.
- Elton, M. (2003). *Daniel Dennett: Reconciling Science and our Self-conception*. Cambridge, UK: Blackwell Publishing Ltd.
- Fodor, J.A: (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, MA: MIT Press.
- Fodor, J.A. (1985). 'Fodor's Guide to Mental Representation: The Intelligent Auntie's Vade-Mecum'. *Mind*, New Series 94(373): 76-100.
- Fodor, J. A. and LePore, Ernest (1992). *Holism: A Shoppers' Guide*. Cambridge: MIT.
- Fodor, J. A. (1996). 'Deconstructing Dennett's Darwin'. *Mind and Language* 11(3): 246-262.
- Haugeland, J. (1982). 'The mother of intention', *Nous* 16: 613-19.
- Hornsby, J. (1997), *Simple Mindedness: In Defense of Naive Naturalism in the Philosophy of Mind*. Cambridge: Harvard University Press.
- Hurley, S. (2003). 'Animal Actions in the Space of Reasons'. *Mind and Language* 18(3): 231 - 257.
- Lycan, W.G. (1981). 'Form, Function and Feel'. *Journal of Philosophy* 78: 24-49.
- Millikan, R. (1984). *Language, Thought and Other Biological Categories*. Cambridge, Mass.: MIT Press
- Minsky, M. (1985). *The Society of Mind*. New York: Simon& Schuster.
- Nozick, R (1981). *Philosophical explanations*. Cambridge, MA: Harvard University Press.
- Quine, W. V. O. (1956). 'Quantifiers and Propositional Attitudes'. *Journal of Philosophy* LIII : 177-86. Repr. in Quine, *The Ways of Paradox*. New York: Columbia University Press.
- Quine, W. V. O. (1960). *Word and Object*. Cambridge, MA: The MIT Press.
- Reichenbach, H. (1938). *Experience and Prediction*. Chicago: University of Chicago Press.
- Rorty, R. (1982). 'Contemporary Philosophy of Mind'. *Synthese* 53: 323-48.
- Rose, N. (1998). 'Controversies in Meme Theory'. *Journal of Memetics - Evolutionary Models of Information Transmission* 2. Available at http://jom-emit.cfpm.org/1998/vol2/rose_n.html.

- Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson and Company.
- Searle, John R. (1983). *Intentionality*. Cambridge: Cambridge University Press.
- Skinner, B. F. (1953). *Science and Human Behaviour*. New York: Macmillan.
- Sellars, W. (1956). 'Empiricism and the Philosophy of Mind', in Feigl, H. and Scriven M. (eds.), *The Foundation of Science and the Concepts of Psychology and Psychoanalysis*. Minnesota Studies in the Philosophy of Science, I. Minneapolis: University of Minnesota Press.
- Sellars, W. (1974). 'Meaning as Function Classification: a Perspective on the Relation of Syntax to Semantics'. *Synthese* 27: 417-38.
- Stich, P. (1981). 'Dennett on Intentional Systems'. *Philosophical Topics* 12: 39-62.
- Stich, P. (1983). 'From Folk Psychology to Cognitive Science'. Cambridge, Mass: MIT Press, Bradford.
- Yu, P. and Fuller, G. (1986). 'A critique of Dennett', *Synthese* 66: 453-76.