Mind Rules

A philosophical essay on psychological rules and the rules of psychology

Ingmar Visser

Supervised by Prof. Dr. Renate Bartsch

Final thesis for Philosophy of Language and Cognitive Science

Department of Philosophy, University of Amsterdam

February 29, 1996

Contents

Chapter 1

Introduction

Mind rules psychology. Throughout psychology rules are used to explain behavior, most notably in cognitive models based on symbol manipulation. Those rules take on different forms and formats but usually it is something like 'if condition c holds, then do action a', where action might be anything from lifting your arm to thinking that it will rain within an hour.[1] With the help of such rules, many theories have been formulated within such diverse fields as psycholinguistics and social psychology. The model of rules seems to be fairly clear. The suggestion is: just think about a computer and you will get the idea. Rules in a computer reside in memory and the central processing unit reads them and follows the instructions which are expressed by them.

The clarity of such a picture of rules in an 'if-then' format is misleading. People are not computers and they certainly do not have central processing units. People simply do not have 'if-then' rules floating around in their heads, waiting to be activated by some condition or other. Or at least, no psychologist, nor a physiologist, has ever convincingly pointed one out to me. Nor am I ever aware of such rules and luckily so. I would go crazy if I were because I would be applying thousands of rules a second, only in the act of typing this document, making all the quite delicate movements with my fingers. Let alone in making up sentences and thinking of the general line of argument to be taken at the same time.

Of course the cognitive psychologist has come up with a solution to this problem. When he speaks of the rules I am following, he means unconscious rules.[2] Apart from the objection that the unconscious is a notably vague and problematic notion, still this does not solve the problem of rules floating around in the head. Cognitive rules are not physical things, one can not find them by cutting up someone's brain, that much is sure. What then are they?

[1] See for example Newell & Simon (1976/1990) for a defence of this view.

[2] Or as-if rules, as I am certain that Dennett would want to call them.

Well, they are functional. The rules which guide our mental life are functional rules. Psychological rules are not physical entities, they are not conscious entities, but they are functional entities. The rules that are frequently alluded to in psychology have a specific function, for example, explaining how people formulate grammatically correct sentences. That is the goal that Chomsky has: explaining how people are able to formulate grammatically correct sentences and how they learn a language as fast as they do. Chomsky is also the one who made symbolic rules popular in psychology, as against the behaviorist scheme of explanation that ruled psychology until the fifties.

In behaviorist psychology people were seen to be determined by reflexes, elicited by stimuli from the environment, much like Pavlov's dogs who started salivating at the sight of a piece of meat. Cognitive psychology gave the mind a life of its own again. The mind rules psychology. It is exactly this independence of mental life, as conceived in psychology, that is problematic. Mental life, for example our linguistic competence, depends very heavily on the people surrounding us. The contents of our mental life depend on our environment, the objects in the world provide us with food for thought. But is it conceivable that we would have the same ability to think without the world being as it is? Do the formal aspects of our mental life also depend on the world as it is? Descartes thought not: the rational soul functions independently of the world; it is just fed with inputs from outside which, for all we know, could be made up by a demon trying to deceive us.

Three questions are central to this essay: what are psychological rules and how do they relate to reflexes? In other words: what does it mean to follow a rule and how dependent or independent is mental life from other behaviors? In answering this question, in the second chapter, I will mainly refer to Wittgenstein's considerations about rule following in his *Philosophical Investigations* and to Chomsky's ideas about rules of language.

The second question revolves around the notion of representation. Given a certain notion of a rule, what kind of model do you need to implement it? What are the criteria to be met by a model, given the properties needed to justifiably ascribe rule following to the model or to a person? The main source for an answer to this second question, to be found in Chapter 3, is Fodor & Pylyshyn's article (1988) on the architecture of cognition. An alternative answer is provided by Smolensky (1988) in terms of connectionist models. In closing the third chapter, I will discuss the 'formality condition' and the principle of autonomy which I think are foundational for cognitive psychology.

Thirdly, I will consider the question of innateness of rules. There seems no doubt that humans are born with an amazing ability of acquiring a language in a short period of time. Chomsky and Fodor claim that this can only happen when the rules of grammar and the concepts of a language are innate. However, it is not very clear what it means for concepts and grammatical rules to be innate. Against the innatist position, one could hold that humans are born with a general purpose learning device that allows one to learn language very quickly. In the fourth chapter I want to discuss this question and ask what it means for a language or a 'language acquisition device' to be innate. In the fifth and final chapter I will get back to these questions and evaluate what the (partial) answers, given in the following chapters, are worth.

Chapter 2

Wittgenstein and the building blocks of psychology

Rules form the backbone of contemporary cognitive psychology, just as reflexes or stimulus-response relations once were the backbone of behaviorist psychology. The notion of a reflex, alternatively called a stimulus-driven association, and the associated mechanisms of classical and operant conditioning formed the core concepts of the first distinct school of experimental psychology at the beginning of the century: behaviorism.[1] From the fifties onwards, the explanatory role of reflex mechanisms was taken over by symbolic rules, analogous to the rules used in the first computers. Rules are now being used to model cognitive processes and behavior throughout such diverse fields as psycholinguistics, developmental psychology and clinical psychology.

Because of the central role rules and reflexes have played and still play in

[1]Strictly speaking, there are some differences between reflexes and stimulus-driven associations although they occur interchangeably in some of the psychological literature. The differences are mainly differences in flavor, rather than meaning. Skinner (1961), the behaviorist *pur sang*, only allows the term reflex in the 'description of behavior'. The term reflex seems to be associated with innate response mechanisms, such as the knee-jerk reflex and the salivation response of Pavlov's dogs (cf. the next section), whereas the term stimulus-driven association is associated with learned or acquired behavior. Except for the innate/acquired distinction, another important distinction is that between spinal and non-spinal reflexes. The reflex mechanism that Descartes described (also see the next section) is considered to describe spinal reflexes, such as the knee-jerk reflex. The afore mentioned salivation response is not a spinal reflex. As can be seen in the case of Pavlov, non-spinal reflexes lend themselves for conditioning, whereas spinal reflexes tend to be less sensitive to conditioning, although they can be suppressed. I will use the terms to denote non-spinal reflexes. This can be done without loss of generality because spinal reflexes can hardly be expected to play any role in cognitive functioning.

psychological explanation, it is important to assess their relative merits and characteristics in general. The main aim of this chapter will be to set rules and reflexes apart in terms of the Wittgensteinian paradigm of rule following behavior. Before doing so, I will briefly discuss the origins and first uses of the notion of a reflex and behaviorist explanation. Then I will review some commentary given on behavioristic explanation, centering around Chomsky's account of linguistic behavior and his conception of rules. Finally I will turn to Wittgenstein's analysis of rule guided behavior and its relation to reflex mechanisms.

2.1 Reflexes and the roots of psychology

Descartes was probably the first person to specify a mechanism of reflexes in an accurate and detailed way. His idea was that muscles are connected with sense organs by little threads which contracted upon stimulation of the sense organs. Skinner describes this as follows:

> "[The thread], acting like a bell rope, opens a valve at a central reservoir, letting the contained fluid flow outward along a pipeline into the muscles, which it activates." (Skinner, 1961, p.323)[2]

Although the reflex offered a completely mechanical way of explaining behavior Descartes, rightly, thought that reflexes could not be all there was to explaining behavior. He therefore posited the existence of the rational soul. The soul could give rise to the same behaviors as the reflex-mechanisms could, the only difference being that it could do so at will, resulting in the essential difference between humans and the beast-machine.

Unbeknownst to Descartes the reflex came to play a major role in experimental psychology, leaving no part for the soul, or any other kind of mental activity, which he thought to be essential.[3] Thus, a major proponent of behaviorist psychology, J.B. Watson, wrote in his *Behaviorism*:

> "The behaviorist advances the view that what the psychologists have hitherto called thoughts is in short nothing but talking to ourselves."

[2]Skinner (1961) gives a short history of the development of the reflex, from Descartes' first description of it, until the modern behaviorist theories. Skinner likens Descartes' description of the reflex to what was later to be identified with spinal reflexes. Besides that, Skinner says that reflexes became so attractive as a means of describing behavior because reflexes could easily be identified with physiological mechanisms, the so-called reflex-arcs.

[3]Malcolm (1977) argued that there is no contradiction between behaviorism and positing the existence of mental states. Just as it is feasible in theories of the physical world to posit intervening variables if needed to make explanation of data easier, more elegant or more plausible. Malcolm mentions, for example, the theory of atoms that was posited to explain some existing experimental data without any allusion to the claim that atoms might be 'real' objects.

(Watson, 1924)[4]

Thinking, the paradigmatic example of mental activity, is to be construed as (subvocally) talking to ourselves. The fact that we do not actually talk to ourselves when thinking is due to 'the influence of social pressure constantly exerted'. In other words: by negative feedback, the actual vocalization, which accompanies thought naturally, is disposed of. Later on Watson mentions conditioning again, in this case conditioning to fire.

> "During the course of many years I become conditioned in a hundred ways to fire. In other words, depending upon the situation I am now in and the series of situations leading up to the present one, I may do one of a hundred things in the presence of fire." (Watson, 1924)[5]

According to Watson then, conditioning is used to explain all there is to human behavior, including verbal behavior. In the course of your life you have learned to discriminatively respond to different situations, mainly through differing negative and positive consequences of your consequent actions on former occassions. The domain of the mental, including thinking, sensations and emotions, ought to be, in Watson's view, completely reduced to the domain of behavior and the stimuli that elicit that behavior, i.e. to reflex-mechanisms.

Pavlov picked up the idea of reflexes and showed how responses could be conditioned to different stimuli, an instance of what was later to be dubbed classical conditioning. At the beginning of the century Pavlov's paradigm of classical conditioning was very influential. Skinner followed up on this with his paradigm of operant conditioning. Using natural variation in behavior he showed how reflexes could be shaped into different kinds of responses, thereby showing how reflexes are the building blocks par excellence in the explanation of behavior. Soon, however, the behaviorist theory ceased to maintain its promising image. Especially in the realm of verbal behavior, it is quite impossible to tell what kind of stimuli elicit the appropriate responses. Psychologists realized that when interesting behavior was concerned, such as linguistic behavior, the relations between stimulus and response were too complex or simply intractable. Verbal behavior is in most cases not elicited at all by specific and detectable stimuli but largely depend on what's going on inside a person. Mental states and mental processes seem to have a life of their own and therefore deserve to be studied independently.[6] Therefore, another approach to explaining the rich mental life of humans was called for in terms of internal and symbolic rules, which guide and control behavior.

[4] Citation taken from an excerpt of "Talking and Thinking", Chapter 10 of *Behaviorism*, to be found in Lycan, 1990, p.15.

[5] Idem, p.21.

[6] I will get back to this at the end of Chapter 3, where I discuss the 'principle of autonomy'.

2.2 Chomskian rules

Chomsky was and still is mostly concerned with verbal bahavior and he more or less introduced the notion of a rule into the field of psycholinguistic research. At the same time, in the fifties, psychology was revolutionized by the invention of the digital computer. Turing's (1950/1990) description of 'computing machinery', in which the main role was played by symbolic rules, gave psychologists an attractive metaphor for human mental functioning. Chomsky sought to define or describe mental activity independently of the behavior that it eventually produces. Whereas the behaviorist generally denied the existence or reality of mental phenomena, Chomskian rules are mental entities of their own right and with their own etiology.

Chomsky argues that the study of language, and of mind in general, is on a par with the biological study of organs such as the heart and the liver:

> "In fact we may think of the study of mental faculties as actually being a study of the body—specifically the brain—conducted at a certain level of abstraction." (Chomsky, 1980, p.2)

According to Chomsky, the level of abstraction is the level of symbolic rules and principles that guide our production and understanding of language. How then can the study of the language faculty, as a 'mental organ', be compared to studying the heart? Certainly, in studying the workings of the heart there is no 'level of abstraction' involved, although in studying the heart one abstracts from microparticles, molecular structures and the functioning of individual cells. The abstraction that Chomsky makes from the detailed workings of the brain to the 'rules and principles of the language faculty' seems to be different from the abstractions involved in studying the heart or the liver. In the case of the heart we can rightly see the heart as a physically continuous unit, whereas that is certainly not the case with the symbolic rules that in Chomsky's view make up the language faculty. After a lot of anatomical research we know that the heart pumps blood through the body. This kind of fact seems to be of a completely different kind than the 'fact' that we have symbolic rules floating around in our head that determine our linguistic behavior. It seems that, in constrast to the case of studying the heart, by cutting up the brains of a few thousand people we are not going to find out how they were able to use language, because we would need our linguistic competence to identify any interesting processes taking place in the brain that can have any bearing on linguistic competence. Note that in the case of the heart and its functioning we are in a different situation: we need not know what functions the heart has prior to investigating the matter. The level of abstraction that Chomsky presupposes in his investigations is not as natural as he seems to think. Especially, it does not follow from saying that a certain level of abstraction is needed in studying the language faculty that the level that he mentions, i.e. that of symbolic rules and principles, is the right level.

Besides the level of abstraction involved, another important element of Chomsky's account of knowledge of language is that he supposes that our knowledge of language is 'a mental state' that is largely determined by genetic factors and only 'fine-tuned' by experience in the early years of childhood (Chomsky, 1980 & 1986). Nobody, I think, would disagree with Chomsky that the innate structure of our brains is one of the determining factors of what we can eventually learn, that there are structures in the brain which constrain the way in which we use and process knowledge. But it is quite another thing to state that the rules of his Universal Grammar are innate *qua rules*. Chomsky (1986) states that he wants to describe the workings of the language faculty, that is: his account should be descriptively adequate to processes in the brain. However, since the rules he posits as processing rules of, for example, constructing grammatically correct sentences are supposed to be unconscious, i.e. principally inaccessible to consciousness, how then could he determine that they are descriptively adequate?

On the other hand, when we view Chomsky's rules of language as product-rules, i.e. specifying the body of correct sentences that we can and do utter, then they are not necessarily internal rules which guide behavior. To make this clear, consider physical laws, such as the law of gravitation. This kind of law also describes regularities in a certain domain of facts, in this case the domain of falling bodies and planetary motion and a lot else. This is not to say that those laws *cause* the falling of bodies or the nearly circular motion of the planets to occur. This example is derived from Searle (1980, p.37) who makes a similar remark as to the rules and principles of language that Chomsky invokes. Although they may correctly predict the behavior of people to a certain extent, for example in grammaticality judgements, that does not by itself imply that those rules or their contents in some way cause or determine that behavior. Chomsky, however, at least suggests that he endorses that position in saying that those rules are largely innate and comprise a certain mental state of people who have mastered a language. Stein (forthcoming) argues that Chomsky's identification of knowledge of language with a mental state, the so-called I-language (internal language), is an incoherent position. The I-language, being a mental state, would be both the object of knowledge, i.e. the language, and the knowledge itself, being the idiolect of the language-user.

Chomsky (1980, p.5) mentions a thought experiment to convince his readers that having knowledge of language, and knowledge in general, consists in having a certain mental state. Think of someone who knows and speaks English like every one of us. Through a tragical accident he looses the capability of speech but somehow the language centers in the brain are themselves undamaged.[7] The speech centers are only cut off from the rest of the brain, the muscles of the tongue and whatever else is involved in actually expressing oneself. After some

[7]Note that such a thought experiment is only conceivable if one assumes a very strong modularity or functional specialization of the brain.

time has elapsed the person miraculously recovers from the injury and thereby retains the powers of speech again, without any learning process taking place. Chomsky suggests that the crucial question now is whether the person involved can be said to have known English in the intervening period. Chomsky takes the question to be an empirical one and says:

> “One can imagine all kinds of evidence that might indicate that the aphasic [...] was in the relevant state—say electrical activity of the brain or evidence from autopsy.” (Chomsky, 1980, p.5)

The answer to the question is thus affirmative. It ‘would seem perverse’ to claim that the person was not in the relevant state in the intervening period and thus knew English, says Chomsky. He then concludes that the physical state of the brain that the aphasic retained during his injuries is rightly called knowledge of language. It is ironical to find that Wittgenstein says something with exactly the opposite import:

> “No supposition seems to me more natural than that there is no process in the brain correlated with associating or thinking [or knowing or speaking a language, we might add]; so that it would be impossible to read off thought-processes from brain-processes.” (Wittgenstein, 1980, p.903)[8]

These quotes from Chomsky and Wittgenstein at least warn us is that we should be careful in imagining all kinds of things without so much as a hint that the imagining has any bearing on reality. Wittgenstein's point should not be taken literally, i.e. of course there are processes in the brain that are correlated with thinking and other mental activities and states, viz. the processes that go on when thinking. The pointe is that the orderliness of linguistic behavior need not necessarily originate from equally orderly brain-processes.

> “But why should the *system* continue further in the direction of the centre? Why should this order [of thought-processes] not proceed, so to speak, out of chaos? [...] Why should there not be a psychological regularity to which *no* physiological regularity corresponds?” (Wittgenstein's emphasis, Wittgenstein, 1980, p.903/904)[9]

Thus, it might turn out to be quite impossible to identify brain-states as corresponding to a certain thought or the ability of speaking a language.

No doubt there are many regularities in linguistic behavior, ‘psychological regularities’, that are worth considering but there is no a priori reason to

[8]Cited in Stern, 1991, p.208.

[9]Also from Stern, 1991, p.208. Note that the terms *correlate* and *correspond* have acquired somewhat different meanings in modern scientific discourse. The term *correlation* is now mainly used as denoting statistical correlation and not so much with a one-to-one correspondence, as Wittgenstein uses the word.

believe that there should be a one-to-one correspondence between these regularities in language behavior and the brain-processes that support or produce the behavior. Wittgenstein, however, argues that even were it the case that some thought-process corresponded to a distinguishable brain-process, that would still be insufficient evidence that the person in question is now following a certain rule, as the cognitivist would have it and Chomsky would certainly agree. In the next section I will discuss how Wittgenstein arrives at that conclusion. Some introductory remarks are in place as to the background of the disagreement.

Chomsky's position is in disagreement with Wittgenstein's identification of meaning and use; the meaning of a word is to be identified with the actual use that people make of that word. Knowing the meaning of a word, according to Chomsky, consists in having the appropriate rule for applying that word to the objects it refers to. Having such a rule in turn comprises being in the right sort of mental state whereas Wittgenstein's position is that the question if someone knows a meaning of a word, or the grammar of a language, is only to be answered by looking at that person's behavior. Does he use that word in the right way? Does she speak in grammatically correct sentences? These are the sorts of questions that need answers if we are to decide if a person knows a language. Chomsky expresses the opposite opinion when saying that

> "[...] behavior is no criterion for knowledge, though it provides evidence for the possession of knowledge, along with much else, in principle." (Chomsky, 1980, p.6)

The essence of Chomsky's account of language is that having knowledge of language is being in the right sort of mental state and mental states are independent (to a certain extent) from behavior, the environment one lives in et cetera. Moreover, that mental state consists of a set of rules and principles that determine or describe the competence of the language-user, but not the performance.[10]

2.3 Wittgensteinian rules

According to Wittgenstein, the most important characteristic of rules of language is normativity, meaning that it is possible to be wrong according to some standard. The meaning of words and the acquisition of the ability to use them are frequently recurring themes in both Wittgenstein's early and later philosophical writings. Central to his *Philosophical Investigations*, henceforth abbreviated as *PI*, is the notion of a rule, invoked in explaining the understanding of the meaning of words, but also in explaining such diverse behavior as mathematical reasoning and the acquisition of skills. The concept of rules which explain

[10] Egbert Fortuin (1996) digs deeper into the question whether Chomsky's rules of language are best viewed as product or process rules, along the lines of Clark's (1990) classificatory criteria for levels of explanation in cognitive science.

behavior turns out to be rather more problematic then is assumed in cognitive psychology.

In his *Philosophical Investigations* Wittgenstein argues against mentalistic theories of language and meaning of words, such as Chomsky's, although Wittgenstein does not of course argue against Chomsky literally. Wittgenstein uses mainly examples from mathematics and word meaning but the arguments also apply to Chomsky's rules of grammar, since they also have a certain normative content as expressed in correctness judgements of sentences.

Rules and normativity are not involved when one of Pavlov's dogs starts salivating at the sight of a piece of meat. Rather, it acts on behalf of an innate reflex. With use of language and other conscious behavior, the case is certainly different. We suppose that the behavior is guided by rational rules which are not mere reflexes, but also contain a normative element: we are obliged to follow the rules on pain of being misunderstood. The fact that not following the rule leads to misunderstanding is most clear in the case of language. To illustrate this, one might imagine what happens when half of the people start using the word *table* to refer to chairs and vice versa. The dog, responding on the basis of his innate reflexes is not in this way obliged to respond in that particular manner, it just does so 'as a matter of brute fact', to use one of Kripke's favorite phrases. It is by this matter of brute fact, the law-like regularity in the dog's behavior, that we conclude that the dog is in fact responding on basis of a reflex. Rephrased: the existence of a reflex is inferred from the fact that the responding behavior is more or less exceptionless.[11]

Wittgenstein invites us to imagine that a mental image, say of a cube, determines how the word *cube* is to be used. This is the kind of mentalistic idea incorporated in Chomsky's theory: one has a rule (or image) which then determines how a word is to be used, or how a sentence is to be formed in order for it to be grammatically correct and appropriate under the given circumstances. Wittgenstein argues that such a mental image, however, can never determine how it is to be used. It is not clear how the mental image relates to the actual object that is denoted by it or how one should use the image to refer to the object.

To illustrate the point, Wittgenstein (*PI*, #137/138) mentions the example of a projection of a cube that results in a pyramid. The projection expresses the way the concept is to be used, and in this case results in the wrong usage. Of course one could postulate yet another mental image or rule which determines the right projection, but that leads to an infinite regress contradicting the fact that our brain comprises a finite resource.[12] The only conclusion to be drawn,

[11]Of course there are complicating factors here; reflex mechanisms might interfere with one another and the like but, as Pavlov showed, under carefully controlled conditions the regularities can be brought out.

[12]The finiteness of our mental resources was and still is one of the reasons to posit the existence of rules, because rules are the means to defining recursive structures and thereby a precondition to infinite productivity which is claimed to be needed to explain our linguistic

according to Wittgenstein, is that a mental image, or for that matter a mental rule, is not sufficient to determine how it is to be applied.

A more transparent example is that of word meanings in a dictionary: a dictionary supposedly gives us a rule to apply a word to certain objects. A dictionary, however, is circular in its explanation of word meanings. Therefore we can never learn a language from a dictionary, since there would be no starting point. It is of course possible to remove the circularity from the dictionary but that would leave us with unexplained words and again it would be impossible to learn the language.[13]

Kripke (1980) mentions yet another example in his exposition of Wittgenstein's considerations which brings out the point most clearly. Suppose I have learned to add two numbers on the basis of a number of examples and an explanation of how to add, as in fact I have. The crucial question according to Kripke is how 'my past intentions regarding addition determine a unique answer for indefinitely many new cases in the future?' (Kripke, 1980, p.8) Suppose further that I have not made any additions with numbers over 56, which is a reasonable supposition since there are always numbers larger than the numbers I have actually used in computing additions. Kripke's sceptic now argues that, confronted with a new arithmetic problem, say $57 + 68$, I would be quite justified in answering 5 instead of 125, as is the correct answer. The reason for this odd idea is that the series of additions I have made in the past does not determine a unique function which yields 125, given the numbers 57 and 68 as inputs. The series of additions I have actually made in the past can be made to accord with an indefinite number of functions. The rationale is the same as in the above examples: mental states, 'past intentions', do not determine a way of rightly adding two numbers, of following a rule in the correct manner. Searle makes a similar point about semantic contents (which can be viewed as rules in the general sense that Wittgenstein advocates):

> "Even given the semantic contents we still have to know what to do with them, how to apply them, and that knowledge cannot consist in further semantic contents without infinite regress." (Searle, 1983, p.152)

The above examples show that mental contents, as conceived by Chomsky and in cognitive psychology in general, do not determine how to proceed in a given case, be it adding or referring to an object or constructing a correct sentence. That means that Chomsky's mental rules will not do as an explanation of the competence of language users; at least not on their own. Wittgenstein expresses the paradoxical result of the argument as follows:

competence. I will get back to the productivity of language in Chapter 3.

[13]Of course there is the possibility of an infinite dictionary in which the two mentioned problems of circularity and the lack of foundations do not occur, but I can not conceive of a way of learning an infinite vocabulary which is not recursively definable.

> "This is our paradox: no course of action could be determined by a rule, because every course of action can be made to accord with the rule. The answer is: if everything can be made out to accord with the rule, then it can also be made out to conflict with it. And so there would be neither accord nor conflict here." (Wittgenstein, *PI*, # 201)

As it stands a rule can not by itself determine the right behavior and the connection between a rule and the behavior it supposedly guides remains unclear. But Wittgenstein goes on to say the following:

> "What this shews is that there is a way of grasping a rule which is *not* an *interpretation*, but which is exhibited in what we call 'obeying the rule' and 'going against it' in actual cases."
> (Wittgenstein, *PI*, #201)

One of the ways to get bootstrapped into 'obeying the rule in actual cases' is being trained or conditioned to do so. Wittgenstein mentions this (possible) connection between a rule and the appropriate behavior in #198.

> "Let me ask this: what has the expression of a rule—say a signpost—got to do with my actions? What sort of connection is there here?—Well perhaps this one: I have been trained to react to this sign in a particular way, and now I do so react to it."(Wittgenstein, *PI*, #198)

At first glance this seems like a straightforward behaviorist resolution, explaining away the rule by a mechanism of conditioning. In the same paragraph Wittgenstein says the following:

> "[A] person goes by a sign-post only in so far as there exists a regular use of sign-posts, a custom." (Wittgenstein, *PI*, #198)

Wittgenstein's point is that following a rule is made possible by the existence of a custom in which the behavior has some kind of use or goal. In the case of the sign-post this is quite clear where the goal is to reach the desired destination.

Wittgenstein's solution of the paradoxical result then consists of two elements. First, one should be trained in the appropriate way to be following a rule and conditioning is one of the possibilities. And, secondly, by this training one is introduced into a custom (or practice or use as Wittgenstein uses the terms interchangeably) which gives rule following its social dimension and explains its normativity. Fogelin summarizes as follows:

> "To learn to follow a rule is to become master of a technique—a technique that is part of a social practice, institution or custom." (Fogelin, 1987, p.162)

In discussing the acquisition of physical skills with the help of rules Searle makes similar remarks as Fogelin and Wittgenstein.

> "The rules do not become 'wired in' as unconscious mental contents, but the repeated experiences create physical capacities, presumably realized as neural pathways, that make the rules simply irrelevant." (Searle, 1983, p.150).

Learning to follow a rule is the same thing as acquiring a technique, a skill, contradicting the cognitivists' view that rules become internalized but still function in the same way as symbolic rules do. The capacities needed to apply rules are anchored in the Background. The Background consists of all kinds of knowledge, or rather know-how, that one is not normally conscious of and that is the precondition to having the mental life we have (Searle, 1983 & 1992).

In addition to practices or customs and training, there is another element in Wittgenstein's analysis of rule following, or rather a precondition to having practices at all. If people would constantly disagree about the most elementary judgements, say colour judgements, a practice such as language would not be possible. Agreement in the most elementary judgements is one of the aspects of our form of life, Wittgenstein says.

> "That is not agreement in opinions but in form of life. [...] If language is to be a means of communication there must be agreement not only in definitions but also (queer as this may sound) in judgements". (Wittgenstein, *PI*, #241/242).

Our life-form also contains reflex-mechanisms and Kripke's 'blind inclinations' to behavior, which are either innate or conditioned from thereon. Searle's notion of Background is rather like Wittgenstein's life-form in that it is a precondition to conscious mental functioning.

If the appropriate training and introduction to a practice are essential to rule following behavior, it follows that Chomsky's rules which are private and innate can not exist, at least not as rules properly conceived of. Moreover, Wittgenstein's considerations point out that rule following behavior is much closer to, and more intertwined with, reflex-like behavior and skill, as witnessed by Fogelin's use of the term technique, than Chomsky would be willing to admit. Remember that Chomsky posited the existence of rules as an *alternative* to the reflexes of behaviorists. The necessity of training or conditioning also guarantees 'the right causal connection' between rule and behavior, which nicely accommodates Searle's objection to Chomsky's position. Remember that Searle's critique on the Chomskian model was that the fact that rules might be descriptive of certain behavior does not imply that those are the rules that are being followed. However, specifying a training process in which the rules are learned, guarantees that those rules actually mediate the behavior so produced.

Chapter 3

The architecture of cognition

After having specified in the previous chapter the building blocks of psychological explanation it is time to move on to the architecture: how rules or reflexes are put together to produce complex behavior. In the late forties of this century two alternatives to behaviorism were developed after many psychologists realized that the behaviorist repertoire was too meagre to account for complex forms of behavior such as language-processing and problem-solving. Hebb (1949) may be viewed as the first modern connectionist[1], whereas Turing's (1950/1990) article,*Computing machinery and intelligence*, contained all the elements of the classical cognitive paradigm which was to rule psychology until the early eighties. For a number of reasons the connectionist approach to modelling cognition did not get much attention[2] till the early eighties, when important new learning-mechanisms were presented in the PDP-books by Rummelhart, McClelland and the PDP-research group (1986).

Turing (1950/1990) outlined the central elements that make up the cognitive architecture: symbolic expressions with constituent structure and structure-sensitive operations, i.e. the operations that are performed depend only on the formal or syntactic structure of the expressions they are operating on and not on their contents or semantics. Another important feature of Turing's approach that is still very much visible in modern theorizing about the mind is the functional distinction between program and computer or in more up to date vocabulary: the distinction between software and hardware. This results in multiple-realizability: a program can be realized, i.e. made executable, on many different computers. The classical cognitive paradigm has an analogous distinc-

[1]See Quinlan (1991) for a short history of connectionist theory and its associationist predecessors.

[2]See, for example, Dreyfus & Dreyfus, 1990 for a review.

tion between mind and brain, where the mind is supposed to be realized in the brain but it could be realized in any other sufficiently complex medium as well. Therefore the mind can and should be studied quite independently of the brain, that is, at a 'certain level of abstraction' (Chomsky, 1980).

Recently connectionist models have become more popular, for example in the fields of learning and pattern recognition, challenging the above sketched picture of the classical computer metaphor of cognition. The reason is that the achievements of symbol manipulation models are very poor on these tasks. The tasks that humans perform unhesitatingly and automatically, such as pattern recognition (letter recognition), have proven to be very hard to model within the classical paradigm. In the following sections I will outline the basic characteristics of both classical and connectionist schemes of representation, drawing mainly on the seminal papers by Fodor & Pylyshyn (1988) and Smolensky (1988), and try to illuminate why the classical model can not explain certain types of mental processes and behavior.

3.1 Classical cognitive architecture

Fodor and Pylyshyn maintain that both mainstream classicists and connectionists are representationalist, i.e. they find that 'postulating mental states is essential to a theory of cognition' and moreover that those states 'function to encode states of the world'. One obvious way then, of attacking their position would be to deny that the representational stance is essential in the explanation of cognitive functioning and the resulting behavior. I will not pursue this line of argument right now but I will do so a little later in the chapter, after having resumed the main points of criticism raised against connectionism by Fodor & Pylyshyn (1988) and the related assessment of the principles of connectionism by Smolensky (1988).

According to Fodor and Pylyshyn the central elements of the cognitive architecture, the representational structure that allows humans to maintain the complex thoughts they have, are a 'combinatorial syntax of mental representations' and a mode of processing that is sensitive to the structure of those representations. Those features are derived from the architecture of the Turing-machine in which symbolic expressions get written on a tape and processing depends only on the structure and not on the content of those expressions. Combinatorial syntax amounts to having primitive symbols, the alphabet of cognition as it were, which are the constituents of complex expressions. Complex expressions or complex mental representations are built by combining primitive symbols by logical connectives. Both primitive and complex representations form the 'language of thought' (Fodor, 1979), the space of (possible) representations.

The second essential commitment of the classical theorist is that mental processes operate on these primitive and complex mental representations on the basis of the constituent-structure of those representations. That means that only the syntactic structure of the expression is needed in deciding the result of the operation and not the semantic content of it. However, the operations being performed on the expressions are to be constructed in such a way that some semantic property of the expressions is held constant. In mathematical proof-theory, from which the idea of this formal approach stems, this property is truth but it could be any semantic property, according to ones wishes; one could for example design meaning preserving operations.

> "In Classical models, the principles by which mental states are transformed, or by which an input selects the corresponding output, are defined over structural properties of mental representations". (Fodor & Pylyshyn, 1988, p.13)

This principle of the formality of representations, i.e. representations should be distinguishable and acted upon only in regard to their formal properties, is dubbed the 'formality condition' by Fodor (1981a, p.226/227). The advantage is that mental operations are definable independent of the context in which they are to be applied. This contradicts the alleged domain specificity of knowledge and skills. It is conceivable that someone is able to produce some argument about a game of chess, but that the same person is incapable of producing an argument that is formally equivalent when the subject of argument is nuclear energy.[3]

Furthermore Fodor and Pylyshyn hold that these properties of the architecture of cognition 'constrain the physical realizations of symbol structures' and that 'the combinatorial structure of a representation is supposed to have a counterpart in structural relations among physical properties of the brain'. This is an important claim since it implies that classicists are committed to saying that their postulated mental representations and their structure are essential causal factors in the production of (new) mental states and behavior.

Their position implies, for example, that a 'part-whole' relationship holds between simple and complex expressions on both the symbolic level and on the level of (physical) realization. In the expression P&Q, P and Q are proper parts of the former expression and because of this processing can be based on the structure of these kind of expressions. This is not the case in connectionist models as Fodor and Pylyshyn reconstruct them, since different nodes represent the expressions P&Q, P and Q respectively. Their reconstruction however concerns only 'ultralocalist'[4] networks and seems not to be completely fair to the broad range of connectionist models available. Connectionist theorists stress

[3]Fodor can and does solve the problem of domain specificity by proposing that the mind is modular. That is, different domains of knowledge have different (physical) locations and accordingly they can have different rules of transformation. It is unclear how many modules one should have to account for the many different capacities and domains of knowledge one has and how the boundaries between them should be drawn.

[4]The term 'ultralocalist' is due to Goschke & Koppelberg's article (1991, p.138) on the representation of concepts in connectionist models.

the importance of distributed schemes of representation that do not allow single nodes to have semantic contents at the conceptual level. In a distributed representation scheme the 'whole-part' relationship can easily be constructed as a relationship between sets and subsets of activated nodes which are to be interpreted as features. I will further explore the notion of distributed representations in later sections.

The classical position outlined by Fodor and Pylyshyn is motivated by well-known arguments which are mainly of linguistic origin. The unbounded or open-ended productivity of the language-user is to be explained by unbounded productivity of thought and that in turn is explained by recursively defined expressions over atomic symbols. A related argument is that of the systematicity of language and hence of thought, i.e. being able to understand a sentence such as 'Joan loves the florist' implies being able to understand 'The florist loves Joan'[5]. From this systematicity of language follows the systematicity of thought because 'understanding a sentence involves entertaining the thought that it expresses' (Fodor & Pylyshyn, 1988, p.39).

According to Fodor and Pylyshyn then, the systematicity and productivity of thought call for a classical architecture with constituent structure expressions and operations that are structure sensitive. Connectionist models fail to capture these properties and can therefore not be proper models of cognition. Fodor and Pylyshyn admit, however, of the possibility that connectionist models provide a useful theory of implementing classical architecture.

3.2 Smolenky and the proper treatment of connectionism

Smolensky sees connectionist models not as mere implementations of classical architecture, as Fodor and Pylyshyn do, but rather construes connectionism as giving explanations at a different level than classical models do. On Smolensky's account of connectionism the level of explanation is intermediate between that of classical models and that of neural models. He dubs this level the subconceptual level. According to Smolensky,

> "[t]he goal of subsymbolic research should not be to replace symbolic cognitive science, but [...] to explain how symbolic computation can emerge out of nonsymbolic computation."
> (Smolensky, 1988, p.22)

[5] I have taken this example from Bechtel & Abrahamsen (1991) rather than the example used by Fodor and Pylyshyn because I liked the footnote Bechtel and Abrahamsen added: "As our small contribution to reducing the gender-typing prevalent in linguistic examples, we have adapted their actual base-example, 'John loves the girl', to one with female grammatical subject and gender-neutral object." (note 2, page 212) I do not see any principled difference between the two sentences but I do approve of the intention expressed in their footnote.

Smolenksy does not want to get rid of symbolic rules altogether but thinks that connectionist models might add something to the classical picture. A similar view is expressed by Hinton, McClelland & Rummelhart (1990), saying that connectionist models are implementations of cognitive rules but that they can also give rise to 'some powerful and unexpected emergent properties'.

Smolensky argues that the main symbolic/conceptual approach to cognition, the classical approach, is the natural route to take since the paradigmatic cases of knowledge and language, being construed as conscious rule interpretation, are usually represented in a formal symbolic fashion. The reason for this explicit symbolic representation is that language and knowledge should be publicly accessible. However, Smolensky's argument goes on to say that it need not be so that the kind of representations found in individuals be necessarily the same as in the public case. Smolensky says:

> "[...] for the purpose of science, it is sound to formalize knowledge in linguistically expressed laws and rules - but it does not follow therefore that knowledge in an individual's mind is best formalized by such rules." (Smolensky, 1988, p.22)

In Chapter 2 the same point came up with respect to Chomskian rules; the fact that certain rules are descriptively adequate of behavior, for example rules that predict that people will give certain grammaticality judgements, does not imply that people are actually guided by those same rules. They might use completely different means to the same end. A similar point is made by Goschke and Koppelberg in relation to the formation of complex concepts by neural networks:

> "Whereas external symbol systems like language may be characterized by syntactic constituency structure, this need not be the case for the internal representations underlying the capacity to use such symbol systems. It may, instead, suffice to assume a coupling of external symbol tokens (e.g., written words) with internal distributed representations of concepts."
> (Goschke & Koppelberg, 1991, p.152)

A lot of conscious behavior is directed by conscious application of rules, e.g. doing mathematics and playing chess. On the other hand there is a lot of behavior that is not directed by conscious application of rules such as 'perception, practiced motor behavior, fluent linguistic behavior, intuition in problem solving—in short, practically all skilled performance' (Smolensky, 1988, p.5). Smolensky introduces the so-called 'intuitive processor' to account for these kinds of behavior and presumably it, the intuitive processor, is best described and explained by a connectionist model. Modelling behavior at the subconceptual level brings with it that, although descriptions at the conceptual level are possible, they are not 'precisely valid'. Only at the subconceptual level accurate computational principles are valid, as opposed to classical models where computation is essential to the conceptual level.

Fodor and Pylyshyn state that

> "[...] systematicity arguments infer the internal structure of mental representations from the patent fact that nobody has a punctate intellectual competence". (Fodor & Pylyshyn, 1988, p.40)

They conclude that connectionist models can not be appropriate models of cognition because they would allow such 'punctate minds', i.e. models of minds that allow 'the ability to think seventy-four unrelated thoughts.' (p.40) However, once again, their construal of connectionist models does not seem to be completely fair. In connectionist models which have a distributed representation-scheme the different states of the network are very much related with each other, resulting in some form of systematicity. Since the state of a network is defined as the activations of its units, which have continuous values[6], neural network states, being the equivalents to mental states, are interrelated in a way that symbolic representations are not. In a distributed neural network it can not be the case that it can be in 'seventy-four unrelated' states, notwithstanding Fodor & Pylyshyn's claim. Their claim is only valid for the 'ultralocalist' network they describe as the paradigm exemplar of neural networks.

3.3 Rules in connectionist models

It is frequently contended that connectionist models can not display systematic behavior because they can not implement rules (cf. Fodor & Pylyshyn, 1988; Levelt, 1989). Putting it this way however, is begging the question against connectionism. Connectionist models do not explicitly use rules in producing behavior and therefore it would eo ipso be impossible for connectionist models to be good models of cognition. In order to get a fair comparison between classical and connectionist models, systematicity should be defined in terms of the tasks to be performed. Garson follows this line of reasoning:

> "However, if we are not to trivialize Fodor and Pylyshyn's Thesis [the thesis that systematicity requires classical architecture], we must clearly distinguish success at syntactic tasks from the use of classical processing. Net systematicity must be defined by the *tasks* that nets must accomplish to be adequate models of cognition." (Garson's emphasis, 1994, p.296)

He then goes on to review some studies in which neural networks are used to model a cognitive task. For example, a network of Servan-Schreiber (1989)

[6]Although in the ideal artificial neural network activations can have continuous values in most simulations that is not the case. This does not really counterfeit the point being made since continuous activations can have arbitrarily close approximations by discrete values.

which was successfully trained to recognise a regular language, i.e. a language that is produced and recognised by a finite automaton.[7] Garson remarks gives the following comment on the workings of the network:

> "[H]ere we see the clear emergence of classical rules in a net trained on a systematic task," (Garson, 1994, p.298)

This is in accordance with Smolensky's remark about the goal of connectionist research, being 'to explain how symbolic computation can emerge out of nonsymbolic computation'. Against this, classicists could hold that this only supports their view that connectionist models only implement classical architecture but that leaves the classicist with the problem to explain how the neural network came to know the rules. Certainly, connectionist models do not implement any explicit rules and during training no connection whatsoever with rules is made. In the next chapter I will elaborate on the function of training and learning in neural networks and on the relation with the possibility of representing rules in such networks. Another study that Garson (1994) mentions is that of Elman (1989) which describes a recurrent neural network that learns to recognise 'subject-verb agreement despite intervening relative clauses.' According to Garson then,

> "[...] this shows at least that some simple forms of context-sensitivity [deemed essential for understanding natural languages] do not pose a serious problem for learning in connectionist nets." (Garson, 1994, p.301)

Networks such as Servan-Schreiber's and Elman's show that connectionist models can model at least some systematic tasks for which Fodor & Pylyshyn would recommend a rule-based classical model. Another example is a network used by Scholtes (1993) to model parsing of natural language sentences which was quite successful. Connectionists stress the importance of distributed representations, which result in such well-known features of connectionist models as graceful degradation and resistance to noisy inputs. Those are characteristics that have proven very difficult to implement within the classical framework of artificial intelligence (Scholtes, 1993; Costall, 1991). Instead of following explicit rules as is usual in the classical model, connectionist models 'satisfy soft constraints' at the subconceptual level (Smolensky, 1988).

Two features of the above sketched examples are important in assessing the main differences between representations in classical and connectionist models. One feature is the need for distributed representations instead of localist representations. I will get back to this later. The second important characteristic

[7] I assume most readers to be familiar with these terms; finite automata and the corresponding regular languages are the simplest languages in the Chomsky-hierarchy of languages and automata of which the Turing-machine and the related recursively denumerable languages form the top elements. See for example Sudkamp (1988) for an introduction to languages and machines.

of a connectionist model of representation is the training-process involved in acquiring representations. This second feature is important because it shows the relation with the Wittgensteinian paradigm of rule-following.

In Wittgenstein's view training is (one of) the way(s) to get introduced to a practice which supports rule-following behavior. The practice is a precondition to the possibility of following a rule and one can only be said to be following a rule when one is a proper member of such a practice. This contradicts however the 'principle of autonomy', the principle that mental functioning and mental states should be studied independently of the semantic properties they have. This principle is taken to be one of the foundational principles of psychological theorizing according to many philosophers (Fodor, 1981; Stich, 1983; See the next section of this chapter for an elaboration of this principle.) The classical model of cognition limits itself to mental states and their interactions and the production of behavior as far as it is possible to specify this at a symbolic or conceptual level.

I promised to get back to the issue of distributed representations. In classical cognitive models, as well as in classical AI-models, representations consist in structured symbolic expressions which reside in an arbitrary place in memory. In connectionist models representations, insofar as one can speak of such, are to be found in the connections between the many nodes the network consists of. The connection strengths assure that an active representation, the activated nodes of the network, is formed in reaction to an input signal. This active representation can thereafter be used to form the output signal again through using connections between nodes. Goschke and Koppelberg, for example, remark the following about concepts:

> "Concepts are transient memory structures that are constructed on the fly [...]." (Goschke and Koppelberg, 1991, p.142)

This illustrates the idea that concepts, or any other symbolic structure, are never permanently available in the mind, but are best viewed as being formed when needed. Moreover, it is represented distributed over many connections and nodes of the network.

One difference then stands out between classical and connectionist models and that is the fact that in classical models all the propositions that it has available are explicitly stored in a separate memory, whereas in connectionist models memory resides in the connection-strengths between units that assure that the right representation is formed. A clear advantage is that the formation of representations is directed by the content of the input. This so-called content-adressability of connectionist models is not achievable in the computational model where representations are stored in a separate memory which is organized along principally different lines. Separate is here meant to refer to the fact that conceptual memory in classical models is separate from the processing rules that transform representations. In connectionist models the connection strengths account for both the activation *and* processing of representations.

Haselager (1995) discusses implicit and explicit representations in connection with the frame problem of AI. Before discussing that problem some remarks on the implicit/explicit distinction should be made. As noted before, in classical models concepts are explicitly stored. There is also a notion of implicitness: concepts that we do not yet know, that we have not yet learned, are implicitly stored, i.e. they are derivable from innate primitive concepts. The notion of implicitness, however, is problematic; only imagine the implicit representations that a blackboard contains. It is possible to draw any symbol or image on a blackboard, representing any arbitrary object. We would not therefore want to say that a blackboard has mind-like capacities of representation. Connectionist models offer something of a middle-ground between explicit and implicit representations. Representations are formed upon presentation of an input. The input is thereby reconstructed (cf. Goschke & Koppelberg, 1991) into a representation. This process of reconstruction is exactly what is missing in the example of the blackboard: it does not add anything to the input, whereas in a connectionist model the representation is automatically related with similar representations.[8]

The frame problem centers around accessibility and relevancy of information or knowledge (Dennett, 1990). In the classical AI-model knowledge is stored in the form of explicitly articulated propositions. New information that is fed into the body of knowledge so stored, can and usually does have consequences for the stored knowledge, i.e. some propositions might have to be changed or abandoned. The discrete propositional form of memory of the classical model, however, brings with it that there is no easy way to decide which knowledge has potentially to be adapted. And more importantly, in order to represent commonsense knowledge one would possibly need millions of propositions, varying from 'gardens usually have trees in them' till 'elections presuppose the existence of a nation-state which is not at this very moment threatened to be swallowed by a Martian giant'. It is quite absurd to suppose that people actually have explicit representations of all such facts. Haselager says that

> "[...] any approach to the frame problem should try to refrain from completely describing the world: it should avoid having to spell it all out". (Haselager, 1995, p.73)

Distributed representations do exactly that: avoiding to have to spell it all out. Instead of having explicit representations, or syntactically tokened representations as Haselager (1995) calls them, neural networks generate an

[8]This explicit/implicit distinction does not exactly mirror the explicit/implicit distinction that is made in psychological literature on memory. In experimental memory research the implicit/explicit distinction is often related to priming-experiments in which subjects have to guess a word that is subliminally presented on a screen. They do better when they are beforehand presented (again subliminally) with related words, for example moon before sun. Implicit in such experiments refers to the fact that subjects are not conscious of the fact that they have seen a related word but still get better scores.

activation-pattern on the basis of past experience and the input which can thereafter be used to generate the appropriate output. Moreover, the changing of connections between nodes implies that all outputs that depend on those connections and nodes, which can be taken to represent features of the input, are changed automatically. Haselager explains in the following way:

> "[...] a distributed representation represents information in a holistic fashion, without its decomposition into constituent concepts and their interrelationships. As the knowledge a system possesses lies embodied in its weights, it directly and automatically constrains the processing of incoming information."
> (Haselager, 1995, p.141)

Connectionist models can be seen to represent knowledge in an intrinsic way, where intrinsic refers to the organisation that is imposed on the different bits of knowledge. In the symbolic model of cognition propositions are seen as completely separate entities and a preferential ordering or organisation has to be imposed on top of them to assure that it doesn't run into problems such as the frame-problem. The relevancy of one hypothesis over another in a certain situation must be explicitly stated in addition to the statement of the proposition itself. Connectionist models automatically have an associative structure that, next to the ability to represent knowledge, ensures that, given a certain input, the relevant propositions are generated. One can rightfully ask whether connectionist models can be called representational at all. When representations are formed 'on the fly' as Goschke & Koppelberg (1991) maintain, one can hardly speak of representations, since the concept of representation implies that the representations are more or less static entities, or at least stable, enduring states. In the connectionist approach representations are only temporary structures that direct behavior and cause other mental processes to occur, rather different from the static representations in the classical model. Stern (1991) argues that we should give up the notion of representation altogether because it implies that we should be able to find neural structures that correspond to them and that has proven impossible.

3.4 The principle of autonomy

Many of the characteristics of the classical architecture of cognition and of the mental rules that Chomsky proposes, derive, I think, from the so-called 'principle of autonomy'. The basic assumption of cognitivist psychology is that mental states should be studied as objects in their own right, as independent and real objects. Chomsky argues that knowledge of language comprises a certain mental state. Although behavior can count as evidence for having this mental state, this need not be so. Even when one does not follow the rules, i.e. when one does not formulate one's sentences according to the rules of grammar, one can still be said to have internalized the correct rules of grammar (Chomsky, 1986). Rules are thus internal and mental. Another important assumption that is implicit in cognitive theorizing about the mind is that rules are supposed to give a complete account of human mental life: nothing else is needed. However, as I argued in Chapter 2, rules do not function on their own. They need a Background, or a basis in a training process and a form of life to be able to function properly.

In section one of this chapter I mentioned the 'formality condition' which is due to Fodor (1981a). It means that mental processes depend on the formal aspects of mental representations. Formal here simply means syntactic. Mental processes thus depend only on syntactic properties of mental representations. The semantic properties of mental states are not seen as playing any role in psychological explanations.

> "Similarly, in specifying that only internal properties and relations are relevant to explanatory psychological properties, the autonomy principle decrees that relations between an organism and its external environment are irrelevant to its current (explanatory) psychological state". (Stich, 1990, p.348)

Fodor takes a similar position, as witnessed by his *Methodological solipsism considered as a research strategy in cognitive psychology*, in which he says that

> "[...] mental states can be (type) distinct only if the representations which constitute their objects are *formally* distinct". (my emphasis, Fodor, 1981a, p.227)

That is, Fodor states that mental states should only be individuated according to their formal properties. The arguments for this position given both by Stich and Fodor go along the lines of Putnam's 'Twin Earth' arguments. Normally, in specifying a mental state one needs a description of the objects that the mental state refers to, because that is what the mental state is about. Naturally, if one thinks that 'water is wet', the thought refers to water. However, if we follow Putnam's thought experiment about a second world which is identical to ours except for the fact that water does not refer to H2O but XYZ, then it would be impossible to say what the thought 'water is wet' is about, because we would not know if the word *water* refers to H2O or to XYZ. Fodor concludes the argument as follows:

> "Surely, however, characterizing the objects of thought is methodologically prior to characterizing the causal chains that link thoughts to their objects. But the theory which characterizes the objects of thought is the theory of *everything*; it's all of science. Hence, the methodological moral of Putnam's analysis seems to be: the naturalistic psychologists will inherit the Earth, but only after everybody else is finished with it. No doubt it's all right to have a research

> strategy that says 'wait awhile'. But who wants to wait *forever?*"
> (Fodor, 1981, p.247/248)

The argument is thus that specifying the world in some objective, scientific vocabulary is essential to relating our mental states to the surrounding world. However, since this process of specification is going to take forever to accomplish we should just forget about semantic properties of mental states and concentrate on the formal properties.

The principle of autonomy, the formality condition and *methodological solipsism* are in fact expressions of the same basic position. From this position follow all the salient features of the classical architecture of cognition: its discreteness, that is the representation of knowledge in discrete propositions, the disconnectedness of mental states from the world, and the formal way of transforming those mental states. Therefore the principle is also responsible for the less advantageous properties of the classical architecture which I have described elsewhere in this chapter.

Chapter 4

Learnability and nativism

In the second chapter of this essay, on rules and reflexes, I referred to Chomsky's views regarding the origin of linguistic knowledge. His view is that our linguistic system, or the language acquisition device, is largely innate and that this must be so because of the poverty of the stimuli that cause it to develop (Chomsky, 1986). Fodor (1981b) has developed a similar view with regard to concept acquisition or 'fixation of belief' as he calls it. Fodor argues that there is no such thing as learning at all in the domain of concept acquisition, but rather that our whole conceptual system is innate. Chomsky argues likewise for the language faculty: it is a fixed, innate structure with predetermined paths of development. As I argued in Chapter 2 it is hard to imagine what such an innate structure as the language faculty should look like. I argued that there is an incompatibility between the innateness of linguistic capacities, expressed as rules, and the fact that language essentially depends on the existence and workings of a community of language-speakers. In other words: there is an incoherence between linguistic capacities being innate and having semantic properties at the same time. This is due to the fact that being innate is something that is determined by genetic, internal factors whereas the semantic properties of a language essentially depend on the bonds between a person, his environment and his peers. Semantic properties at least involve the relation between a human and his surroundings whereas innate properties, by definition, do not depend on such relations. Essentially the Wittgensteinian criticism on Chomsky boils down to saying that one cannot justifiably ascribe rules to an individual person but rather that one needs to consider the practice in which that person participates in as well. Besides that, I mentioned the training process that would be needed to acquire rule-following behavior. This last aspect of rule-following behavior is the aspect I want to consider in this chapter. I want to review some of the authors that have taken position in the debate about innatism and see what role is left for learning in the acquisition of rules.

4.1 Learning as induction

Learning concepts and grammatical constructions in your native language can be viewed as the acquisition of rules. In the case of concepts one learns the rule that selects the defining properties of the objects that the concept refers to and one thereby learns to apply the concept or word to the appropriate (class of) objects. In the case of grammar the rules to construct proper sentences and to analyze, i.e. understand, incoming sentences, are learned. In the empiricist tradition this learning process is described as a process of induction. Based on experience one learns to use a concept, for example by a process of ostensive definition. The presentation of a suitable number of instances of a concept finally allows for a generalization to be made: the induction is completed.

In Fodor's (1981b) reconstruction of the Empiricist approach to concept-acquisition, sensory experiences form the backbone of the concepts acquired by an inductive process; sensory experiences directly trigger primitive concepts and complex concepts are construed from the primitive concepts by some combinatorial apparatus. Fodor describes the process of concept-acquisition as follows:

> "The theory is this: concept learning involves the inductive fixation of certain beliefs, and the mechanisms of concept learning are realizations of some species of inductive logic. In particular, they involve the formulation and confirmation of hypotheses about the identity of the concept being learned." (Fodor, 1981b, p.267)

There are two ingredients in this explanation of concept-acquisition: learning is induction and secondly, more specifically, induction is the testing of hypotheses. To be able to do inductive reasoning, Fodor mentions four conditions that have to be fulfilled:

1. A source of hypotheses to be tested; which in turn presupposes a canonical format and vocabulary for the specification of the hypotheses.

2. A canonical format and vocabulary for the representation of the data.

3. An a priori ranking of the hypotheses in order of relative simplicity. (Roughly because, barring such a ranking, a body of data bestows equal confirmation on all of the infinitely many non-equivalent hypotheses with which it is formally compatible.)

4. A confirmation metric: roughly a function from pairs consisting of hypotheses and data sets onto numbers that express the level of confirmation of the hypothesis relative to the data. (Fodor, 1981b, p.268/269)

The argument from these conditions to nativism is easily made: testing hypotheses presupposes the concepts in which the hypotheses are expressed, the 'canonical format' (which is simply the 'language of thought') of the second condition. Hence, the 'source of the hypotheses to be tested', the mechanism that produces hypotheses, and the 'canonical format', must be innate because no suitable alternative is at hand (Fodor, 1981b, p.270). Moreover, it follows that only complex concepts can be learned, i.e. concepts that are built out of primitive concepts with the help of some combinatorial apparatus, which, of course, should be innate as well. Which concepts are to count as primitive is debatable, but that does not really affect the line of the argument.[1]

In the classical model then, using or entertaining concepts, or for that matter using any symbolic forms, is seen to stem from innate capacities. The most important argument for this position is, as mentioned before, the poverty of the stimuli on the basis of which humans learn concepts and language. We only need very little experience to be able to use a concept. Fodor expresses the importance of this to the Rationalist tradition as follows:

> "[...] it's the Rationalist view that the normal relation between lexical concepts and their occasioning experiences is brute-causal, i.e. 'merely' empirical: such experiences function as the innately specified trigger of the concepts which they—to borrow the ethological jargon—'release'". (Fodor, 1981b, p.280)

Fodor then discusses an example in ethology to illustrate that very little stimulation is needed to evoke certain behaviors. A newborn duck will treat the first moving object it sees as its 'mother' and start following it. Humans acquire concepts likewise: a single presentation of a certain object is supposedly enough to 'trigger' the concept and make it available for further usage. According to Fodor this can only be explained when the concepts are innate. Chomsky (1986) gives a similar argument, essentially stressing the poverty of stimulus, to prove that the rules of language, and more specifically the rules of grammar, are innate.

4.2 The learning-paradox

Fodor's (1981b) view on learning concepts naturally leads to the somewhat paradoxical conclusion that learning is in fact impossible. When learning is seen as induction, as Fodor and Chomsky do, and when learning is in turn construed as a process of hypothesis formation and confirmation, then the 'source of the hypotheses' must be innate. The 'source of the hypotheses' consists, in the case of concepts, of primitive concepts and a combinatorial apparatus. It easily follows

[1] According to Fodor (1981b), the only disagreement between the Empiricist and Rationalist traditions is the class of concepts they take to be primitive. Fodor takes all lexical concepts to be primitive, that is all concepts that are expressible in one word (with a few ifs and buts added). His argument for this position goes something like this: lexical concepts have proven to be unanalyzable and therefore they must be primitive, because otherwise one could analyze them into the primitive concepts they are built from. As a consequence of Fodor's position the country in which you are born determines what concepts you have available innately, since what is and what is not a lexical concept is different from one language to another.

from here that all concepts that one will ever acquire are available from birth. Piatelli-Palmarini confirms this attitude of Fodor's towards learning concepts:

> "Nothing new, he [Fodor] contends, can be 'acquired' during cognitive development, and this for purely logical reasons."
> (Piatelli-Palmarini, 1980, p.143)

The paradox consists in the conclusion that learning is in fact impossible and that all concepts are innately specified. This construal of Fodor's position shows it to be somewhat empty, because it mainly amounts to saying that the state in which we are born is a precondition to whatever we learn during the course of our lives. Well, that is certainly true and I can not think of anyone who would want to disagree; whatever language we learn, however many concepts we acquire during a lifetime, being born is certainly a precondition for it to happen. What needs explaining however, is how the interaction with the environment 'triggers' the concepts that we in fact come to acquire. Merely saying that concepts are innate and that they are triggered by this or that is not enough. We need a specification of what it means for a concept to be innate and a process which specifies what interaction with the environment is needed to actualize the concept in question. In Fodor's model of concept-acquisition that interaction consists of 'triggering' innate primitive concepts which are then combined in a logical way; complex concepts thus consist of conjunctions and disjunctions of primitive concepts or features. This combining of primitive concepts in conjunction and disjunction is the only way of acquiring new concepts.

One proviso should at least be made to Fodor's analysis of the learning-process: it is only valid when induction is seen as hypothesis-testing. A logical place then to confront Fodor, is on this point: are there any alternatives to hypothesis-testing as a mode of learning? Behaviorist learning mechanisms do not at first sight classify as a 'species of inductive logic' and hypothesis-testing. The behaviorist representation-scheme is, however, very impoverished. In fact behaviorism does not have a notion of representation at all; behaviorism only allows pairings of stimulus and response *without* any intervening processes. The cognitivist approach, so to speak, *only* allows 'intervening processes', leaving open the question between what and what they intervene (Fodor, 1981a), because it has proven difficult or even impossible to specify stimuli in such a way as not to make it question-begging with respect to the expected response.

Molenaar (1986) argues that the formal approach inherent to functionalist (symbolic, classical, cognitive) psychology disregards the implementational properties of actual cognitive systems, such as non-linearity, gradual levels of activation and parallelism. From the domain of information science we know that formal, symbolic systems have fixed rules of transformation. By definition such systems are built with a fixed alphabet of primitive symbols (primitive concepts) and a predetermined set of logical connectives. With the fixed alphabet and the logical connectives complex expressions (complex concepts) are constructed. Transforming those expressions into new expressions is done with the help of a preset stack of transformation rules. These are, loosely put, the defining properties of formal systems. If we take the mind to be a formal system, as Fodor does, the implementational properties are excluded from the model. Molenaar (1986) mentions for instance the property of nonlinearity of artificial neural networks (and neurons in real brains) which may have a profound influence on the functioning of the brain. Another important property of real-life neurons and nodes in neural networks is the real valued activation that they can have, resulting in the possibility to account for gradual differences between features of given inputs. Molenaar says the following about formal cognitive systems as they are studied by cognitive science:

> "It is a key feature of such disembodied systems that they can only operate according to explicitly implemented rules of information integration. Accordingly, formal information systems have a knowledge base that is invariant save for transformations induced by these explicit rules". (Molenaar, 1986, p.247)

The strong innatist position that Fodor takes can be seen to follow from the formal-systems approach (the 'formality condition') that he advocates for studying cognitive systems. Once we allow implementational properties, i.e. properties of the hardware of the mind, to play a role in explanation, the innatist position is not so certain anymore. It is important to note that on the one hand the funcionalist approach to mental phenomena disregards matters of implementation and on the other hand it also disregards, in principle, the semantic properties of mental states, as I argued in the last section of Chapter 3.

4.3 Alternative hypotheses

Fodor's argumentation that concepts must be innate is very tight. When leaving that argument as it is, another problem poses itself and that is that Fodor's position seems somewhat empty. It doesn't clarify in any way what it means for concepts to be innate. The same problem is inherent in the innate linguistic structures that are proposed by Chomsky; the way in which he describes them is not in any way suggestive of how those structures might be realized in the brain.

The model of learning proposed by the behaviorists is intuitively plausible and does not classify as inductionist learning. Given the fact that 'learning', as Fodor describes it, does not classify as learning at all, the behaviorist paradigm of conditioning counts as the only mode of learning that is available. Moreover, as I argued in Chapter 2, conditioning can well be viewed as a manner of acquiring rules or at least as a precondition to the possibility of acquiring behavior that is rightfully ascribed the predicate *rule-guided*. So it seems that Fodor's position that learning requires an innate 'source of hypotheses' which is structured like a language is too strong. Associative, behaviorist learning qualifies as a candidate for an alternative to learning by induction. The behaviorist model however, lacks a scheme of representation that is rich enough to model the intricate mental processes of our species. Boom summarizes the cognitivist and the behaviorist positions as follows:

> "Nativism, which is derived from rationalism, denies all novelty (structure without genesis), while behaviorism, which is derived from empiricism, denies the importance of previous structuring (genesis without structure)." (Boom, 1991, p.274/275)

The rationalist tradition leads unavoidably to nativism about concepts and other important functions of the mind such as language learning. Behaviorism, stemming from the empiricist tradition, misses out on the mental life of people altogether, because it denies the autonomous workings of the mind. What is needed on top of the behaviorist theory of learning is the possibility to make use of an extensive scheme of representation such as cognitive models have. Put the other way around: what is needed on top of the representational structure of cognitivism is a way of acquiring genuinely new concepts and behaviors.

Connectionist models provide both 'structure' (representational capacity) and the possibility of 'genesis' (learning). The representational structure of a neural network consists in the nodes and connections between them. The possibility of 'genesis' of a neural network lies in the learning mechanisms that have been developed to 'train' networks. This in contrast to classical process models in psychology which, if they are ever implemented at all, are, so to speak, handmade, preprogrammed devices with explicit symbolic representations and rules of transformation. I will now give a short description of the way that connectionist networks are usually trained and then turn to the question in what way those networks can be seen to represent concepts.

Connectionist models start out with randomly set strengths of connections between nodes. Then the input is presented on the designated input-nodes of the network and the processing begins. The input nodes pass on their activation to one or several hidden layers of units and from those nodes in turn activation spreads to the output-nodes. In the learning process that follows the weights or connection strengths are changed on the basis of feedback to the network. The feedback consists of an error value that is based on the difference between the actual output and the desired output. This process is repeated until some criterion of functioning is reached, usually something like the error value(s) being smaller than some preset value.[2] This description shows the connectionist approach to learning to be thoroughly behaviorist: learning is based on feedback from outside. The connectionist approach adds to the behaviorist model a powerful representational structure. Bereiter (1985, p.208) describes this model

[2]See for example Quinlan (1991) for an overview of architectures, learning procedures, et cetera of connectionist models.

of learning, based on feedback, as 'chance plus selection': the organism gets the right response by chance and thereafter selection by feedback assures that similar responses are produced on similar inputs. In fact, this is an instance of operant conditioning in which the desired response-behavior is reinforced to make sure that it will occur again on similar occasions.[3]

The representational structure of connectionist models does not use a classical symbolic scheme of representation but rather a subsymbolic scheme (Smolensky, 1988; Hofstadter, 1985). As can be seen from the above given description, processing in connectionist models depends on activation values of nodes and the connection strengths between nodes. What is actually processed then, are vectors or arrays of activation values, each value representing the partial absence or presence of a (micro-)feature of the object represented. The (micro-)features are the subsymbols. This is radically different from the classical picture where features or primitive concepts are either present or absent. In the classical account of (complex) concepts, concepts are sets of primitive concepts (features) which are combined by logical connectives which does not allow partial presence of absence of features. Moreover, in connectionist models more intricate relations between features are possible. For example, the presence of a certain feature can enhance or diminish the activation of another feature; a possibility that is very hard to implement in a formal symbolical system.

The representational capacity of a neural network resides in the connections between nodes and the weights or strengths that are attached to these connections. Concepts in neural networks are usually said to form in the hidden layers of the network, that is the layers of nodes that are neither input- nor output-nodes (Elman, 1990). The activation of the hidden layers of a neural network results from the combined activations of the nodes from previous layers, allowing for complex combinations of present and absent features. Moreover, the representations are formed by feedback from outside the network, assuring that they are directly linked to appropriate outputs of the network.

This last characteristic of connectionist models conforms to the theory that Campbell & Bickhard (1987) have developed of so-called interactive representations. A (cognitive) system is seen to represent something when it can interact with it in the appropriate way, which does away with the static notion of representation that prevails in the classical model.

> "To know something interactively is to interact with it successfully according to some goal." (Campbell & Bickhard, 1987, p.54)

And later they add the following to this:

> "Learning constructs new system organization that succeeds in interacting with the environment, and in differentiating that environment in useful ways." (Campbell & Bickhard, 1987, p.55)

[3]It should be noted that other learning mechanisms exist which are not based on feedback but only on strengthening of associations between directly linked nodes of the model.

According to Campbell & Bickhard then, the most important ingredient of representation is that it should allow a system to interact with its environment in the appropriate way.

Connectionist models can hardly be said to have concepts available from scratch as do classical models. The training process is essential for the network's acquisition of representations. Of course, as a precondition to learning, a neural network should have basic discriminative powers, as humans have. Without the possibility of discriminating red and green, for example, one could hardly be expected to be able to learn the concepts *red* and *green*. It is inherent in the classical position that from this ability to discriminate red and green follows that one has the concepts *red* and *green* available. In the connectionist model this is not so. Although there should be, on the input layers of the model, nodes that can discriminate between red and green, that does not imply that the model has the concepts available. The network still has to learn to respond in appropriate ways to those colors before the model can be said to have acquired the concepts. This differentiating capacity of neural networks provides a notion of innateness that is not as far-fetched as in the classical model. Bartsch (1994) describes these basic discriminative powers as similarity spaces that are available from the start, i.e. before the learning phase begins. They can be likened to the reflexes of Chapter 2 that were a precondition to learning rules. There I argued, in agreement with Wittgenstein, that a prerequisite to learning rules was a certain 'form of life' which in part consists of reflexes. The connectionist model, I think, gives a clearer understanding of what these reflexes could be: the ability to discriminate certain stimuli. At least internally, on one of the hidden layers of the network, the different stimuli should result in different responses. The final response, on the output layer of the network, can be formed through conditioning of these reflexes.

Chapter 5

Summary and conclusion

In the introduction to this essay I had three questions; it is time now to evaluate the (partial) answers I tried to provide in the three preceding chapters. The questions were:

- What exactly does the notion of a rule, that is so frequently used in psychological explanations, amount to and what is its relation to the notion of a reflex that was popular in behaviorist psychology until the forties?
- How are aforementioned rules represented in the different models of cognition that have prevailed in modern psychology?
- How do the different modes of representation relate to questions of learning and nativism?

In the second chapter I concentrated on the conception of rules found in Chomsky's writings because he was the first to use rules in the formal sense that prevails in modern psychological explanation. Chomskian rules were seen to be formal, syntactic entities that describe how certain mental representations are transformed into others. In effect, they specify the so-called competence of the language user, i.e. the rules he or she uses to construct grammatically correct sentences. It remains unclear however if Chomsky's rules ought to be viewed as processing rules, that is rules that specify the process by which sentences are produced, or rather that the rules specify the outcome of such a process which needs to be specified further. For the line of argument however, it does not really matter. In the cognitivist framework, of which Chomsky is an important proponent, such a process would be specified in the same symbolic, mental rules as he invokes in explaining the workings of the language faculty.

Along the lines of Wittgenstein's critique of rule following, I criticized Chomsky's account of linguistic competence. The core of the critique is directed at the autonomy of the rules posited by cognitive psychologists in general and especially by Chomsky. Rules and representations within the cognitive framework are seen as internal, mental entities. I argued that Wittgenstein's critique prohibits symbolic mental states to have the semantic properties that are associated with mental states as ordinarily conceived. A symbolic mental image does not by itself determine what it is an image of, i.e. it does not determine to what it refers. The same holds for rules, such as used in mathematics: a symbolic rule does not guarantee that one knows what to do with it, or how to 'proceed in new cases' as Wittgenstein puts it. Searle has similar remarks about the workings of semantic contents: they do not function on their own. Rather, a background of practices and skills is a prerequisite to their functioning (Searle, 1983). The central problem then with rules is that we expect them to have semantic properties, that is properties that relate them to objects outside of the mind that is said to have those rules. These semantic properties are not inherent to symbolic rules. Wittgenstein argues that rule following can only be ascribed to individuals who are embedded in a social practice. In the case of language this is evident: one can only be said to have linguistic rules within a community of people who share the same linguistic uses. Without this community the use of symbolic entities, such as words and sentences, would be quite useless.

Besides this, Wittgenstein provides us with a picture of the possibility of acquiring rules on the basis of conditioning. A precondition to this process of conditioning taking place is that one must have certain (innate) reflexes and discriminative powers. Without the possibility of discriminating red and green, one could hardly be expected to be able to learn the concepts *red* and *green*. The genetic endowment that is the precondition to learning language (and probably much else) is part of what Wittgenstein calls our 'form of life'. It must be noted though that the notion 'form of life' remains pretty vague throughout the *Philosophical Investigations*. The conclusion to be drawn from the second chapter is that Chosmkian rules lack this embedding within a social context and the links with skills and other physical capacities to function properly.

The core of Chapter 3 is the representational structure that is used in classical models of cognition. Fodor & Pylyshyn (1988) maintain that mental representation is best viewed as a language-like phenomenon, consisting of basic symbols (the 'alphabet') and a set of rules to combine those symbols into more complex structures. Those rules are best compared to the rules that are used to construct formal languages: logical connectives and the like. Moreover, transforming mental states, i.e. complex symbolic structures, depends only on the formal aspects of such states. The formal approach inherent in cognitivism makes for a range of problems with the classical architecture, as I argued in Chapter 3. Among those problems is the frame problem. When representation is strictly constructed as discrete and propositional, there is no 'natural' ordering of propositions: it is hard to tell which propositions are relevant to a certain problem that is presented. I argued that distributed representations, such as found in connectionist models can overcome this problem, because they use a way of processing that is sensitive to the content of incoming signals.

Fodor & Pylyshyn (1988) state that connectionist models can not be used to model complicated domains, such as language, because modelling such a domain calls for representations with constituent structure and rules that are sensitive to those structures. With the help of a few examples, supplied by Garson (1994) and Elman (1989), I tried to convince the reader that connectionist models can be used to represent some systematic tasks. Although these tasks might be too simple to be compared to the difficulties that are involved in using natural language, at least connectionist models can incorporate some of the features needed for processing language. At the end of the third chapter I mentioned the principle of autonomy from which the most important features of the classical architecture derive.

In the fourth chapter I discussed learning. Fodor (1981b) reconstructs learning as the inductive confirmation of hypotheses. The 'source of the hypotheses' must therefore be innate. This characteristic of models with classical architecture is expressed by the learning paradox: nothing new can be acquired through learning. All the knowledge that such a model can obtain is already (implicitly) given in the initial structure, the 'source of the hypotheses', the 'language of thought'. The only real model of learning that exists then, is the behaviorist model of classical and operant conditioning. Learning through conditioning is one of the possibilities of overcoming the learning-paradox. The behaviorist model however does not have a representational structure needed to accomplish that. I argued that connectionist models combine a representational structure with behaviorist learning capacities. Connectionist models exhibit learning, without presupposing capacities of representation that are prestructured like the 'language of thought' is. To be able to learn new concepts or new rules, the only thing that is presupposed in connectionist models is a discriminative ability, meaning that it must be able to discriminate between the objects or symbolic structures of which it will acquire representations. Finally, representations in connectionist models are best viewed as the ability to discriminatively respond to the things that are supposed to be represented. This ensures a strong connection between representation and action; certainly one of Wittgenstein's demands for mental life to be meaningful, as witnessed by his identification of meaning and use of words and language. The ability to learn and overcome the learning paradox provides, I think, the best reason to favor connectionist models over classical ones with their symbolic rules.

There is more to mind, and hence to psychology, than formality, notwithstanding Fodor's claim to the contrary. Content should also play a part, although it is far from clear what part that should be. Connectionism should get the deserved chance to fail, as did cognitivism, is the reaction of Dreyfus & Dreyfus (1990) to the rise of connectionism. Mind rules psychology, and so it should be, it is its sole subject. However, it does not imply that mind rules should rule psychology.

Bibliography

[Bartsch (1994)] Bartsch, R. (1994), The relationship between connectionist models and a dynamic data-oriented theory of concept formation (Research report No. LP-94-19), Amsterdam: University of Amsterdam, Institute for Logic, Language and Computation

[Bech&Abr (1991)] Bechtel, W. & Abrahamsen, A. (1991), Are rules and symbols needed?, in: *Connectionism and the Mind: Introduction to parallel processing in networks*, Oxford: Blackwell pp.210–254

[Bereiter (1985)] Bereiter, C. (1985), Toward a solution of the learning paradox, *Review of Educational Research, 55*, pp.201–226

[Boom (1991)] Boom, J. (1991), Collective development and the learning paradox, *Human Development, 34*, pp.273–287

[Ca.&Bi. (1987)] Campbell, R.L. &Bickhard, M.H. (1987), A deconstruction of Fodor's anticonstructivism, *Human Development, 30*, pp.48–59

[Chomsky (1980)] Chomsky, N. (1980), Rules and representations, *The behavioral and brain sciences, 3*, pp.1–61

[Chomsky (1986)] Chomsky, N. (1986), *Knowledge of Language: Its nature, Origin and Use*, Convergence, London: Praeger

[Clark (1990)] Clark, A. (1990), Connectionism, competence and explanation, in: Boden, M.A. (Ed.) (1990), *The Philosophy of Artificial Intelligence*, pp. 281–308

[Costall (1991)] Costall, A. (1991), Graceful degradation: Cognitivsm and the metaphors of the computer, in: Still, A. & Costall, A. (Eds.), *Against Cognitivism: Alternative Foundations for Cognitive Psychology*, NY: Harvester Wheatsheaf, pp.151–169

[Dennett (1990)] Dennett, D.C. (1990), Cognitive wheels: The frame problem of AI, reprinted in: Boden, M.A. (Ed.) (1990), *The Philosophy of Artificial Intelligence*, pp. 147–170

[Dre&Dre (1988)] Dreyfus, H.L. & Dreyfus, S.E. (1988/1990), Making a mind versus modelling the brain: Artificial intelligence back at a branch-point, reprinted in: Boden, M.A. (Ed.) (1990), *The Philosophy of Artificial Intelligence*, pp. 309–333

[Elman (1990)] Elman, J. L. (1990), Finding structure in time, *Cognitive Science, 14*, pp.179–211

[Elman (1993)] Elman, J. L. (1993), Learning and development in neural networks: the importance of starting small, *Cognition, 48*, pp.71–99

[Fodor (1979)] Fodor, J.A. (1979), *The language of Thought*, Cambridge, Massachusetts: Harvard UP

[Fodor (1980)] Fodor, J. A. (1980), Fixation of belief and concept acquisition, in: Piatelli-Palmarini (Ed.), *Language and learning: the debate between Jean Piaget and Noam Chomsky*, London: Routledge & Kegan Paul

[Fodor (1981a)] Fodor, J.A. (1981), Methodological solipsism considered as a research strategy in cognitive psychology, in: *Representations*, Brighton (Sussex): The Harvester Press, pp.225–253

[Fodor (1981b)] Fodor, J.A. (1981), The present status of the innateness controversy, in: *Representations*, Brighton (Sussex): The Harvester Press, pp.257–316

[Fod&Pyl (1988)] Fodor, J.A. & Pylyshyn, Z.W. (1988), Connectionism and cognitive architecture: A critical analysis, *Cognition, 28*, pp.3–71

[Fogelin (1987)] Fogelin, R.J. (1987), *Wittgenstein* (Sec. ed.), London, New York: Routledge & Kegan Paul Ltd

[Fortuin (1996)] Fortuin, E. (1996), *De status van taalregels bij Chomsky*, Doctoral thesis (Dutch) from the Department of Philosophy, University of Amsterdam

[Garson (1994)] Garson, J.W. (1994), Cognition without classical architecture, *Synthese, 100*, pp.291–305

[Gos&Kop (1991)] Goschke, T. & Koppelberg, D. (1991), The concept of representation and the representation of concepts in connectionist models, in: Ramsey, W. Stich, St. & Rummelhart, D.E. (Eds.), *Philosophy and connectionist theory*, Hillsdale, NJ: Lawrence Erlbaum

[Haselager (1995)] Haselager, W.F.G. (1995), *The right frame of mind: Cognitive science, folk psychology and the frame problem*, PhD-thesis from the Vrije Universiteit (Amsterdam)

[Hin&al. (1986)] Hinton, G.E., McClelland, J.L. & Rummelhart, D.E. (1986), Distributed representations, reprinted in: Boden, M.A. (Ed.) (1990), *The Philosophy of Artificial Intelligence*, pp. 248–280

[Hofst. (1985)] Hofstadter, D. (1985), Waking up from a Boolean Dream, in: *Metamagical Themas*, London: Bantam Books

[Kripke (1982)] Kripke, S. (1982), *Wittgenstein on rules and private language: an elementary exposition*, Oxford: Basil Blackwell.

[Malcolm (1977)] Malcolm, N. (1977), Behaviorism as a philosophy of psychology, in: *Thought and Knowledge*, Ithaca and London: Cornell UP, pp.85–103

[Molenaar (1986)] Molenaar, P. C. M. (1986), On the impossibility of acquiring more powerful structures: a neglected alternative, *Human Development, 29*, pp.245–251

[Piatelli (1980)] Piatelli-Palmarini, M. (Ed.) (1980), *Language and learning: the debate between Jean Piaget and Noam Chomsky*, London: Routledge & Kegan Paul

[Piatelli (1989)] Piatelli-Palmarini, M. (1989), Evolution, selection and cognition: From 'learning' to parameter setting in biology and the study of language, *Cognition, 31*, pp.1–44

[Quartz (1993)] Quartz, S.R. (1993), Neural networks, nativism and the plausibility of constructivism, *Cognition, 48*, pp.223–242

[Quinlan (1991)] Quinlan, P. (1991), *Connectionism and psychology: a psychological perspective on new connectionist research*, Hertfordshire: Harvester-Wheatsheaf

[Rum.&al. (1986)] Rummelhart, D.E. & McClelland, J.E. (Eds.) (1986), *Parallel Distributed Processing*, Cambridge (Mass.): MIT Press

[Scholtes (1993)] Scholtes, J. C. (1993), *Neural networks in natural language processing and information retrieval*, PhD-thesis from the University of Amsterdam, Department of Computational Linguistics

[Searle (1980)] Searle, J.R. (1980), Rules and causation (Commentary on Chomsky (1980): Rules and representations), *The behavioral and brain sciences, 3*, pp.37–38

[Searle (1983)] Searle, J.R. (1983), *Intentionality*, Cambridge (Mass.): Cambridge UP

[Searle (1992)] Searle, J.R. (1992), *The Rediscovery of the Mind*, Cambridge (Massachusetts) & London (England): The Mit Press

[Smolensky (1988)] Smolensky, P. (1988), On the proper treatment of connectionism, *Behavioral and brain sciences, 11*, pp.1–74

[Skinner (1961)] Skinner, B.F. (1961), The concept of the reflex in the description of behavior, in: *Cumulative Record*, New York: Appleton Century

[Stein (forthc.)] Stein, H. (forthcoming), *The Fiber and the Fabric*, PhD.-thesis from the ILLC, University of Amsterdam

[Stern (1991)] Stern, D.G. (1991), Models of memory: Wittgenstein and cognitive science, *Philosophical psychology, 4*, pp.203–218

[Stich (1990a)] Stich, St. (1990), Autonomous psychology and the belief-desire thesis, in: Lycan, W.G. (Ed.), *Mind and Cognition*, Oxford: Basil Blackwell, pp.345–361

[Stich (1990b] Stich, St. (1990), An excerpt from "The syntactic theory of mind", chapter 8 of *From Folk Psychology to Cognitive Science*, in: Lycan, W.G. (Ed.), *Mind and Cognition*, Oxford: Basil Blackwell, pp.361–371

[Turing (1950)] Turing, A.M. (1950), Computing machinery and intelligence, reprinted in: Boden, M.A. (Ed.) (1990), *The Philosophy of Artificial Intelligence*, pp. 40–66

[Watson (1924)] Watson, J.B. (1924), An excerpt from "Talking and Thinking", chapter 10 of *Behaviorism*, in: Lycan, W.G. (Ed.), *Mind and Cognition*, Oxford: Basil Blackwell, pp.14–22

[Wittgenstein, PI] Wittgenstein, L. W. (1978),*Philosophical Investigations*, Oxford: Basil Blackwell (Standard English translation by G.E.M. Anscombe)