



Updating Empiricist Mentalist Semantics

Barry Nouwt
Master's Thesis Cognitive Artificial Intelligence
Department of Philosophy, Faculty of Arts and the Humanities
University of Utrecht
Supervised by Dr. H.L.W. Hendriks

June 2008

Preface

During my computer science education I learned about the goals of artificial intelligence; its interdisciplinary approach complemented my wish to get a broad overview of important scientific subjects. Accordingly, I started studying Cognitive Artificial Intelligence at Utrecht University after obtaining my bachelor's degree in computer science, and this really stimulated my interest in science and philosophy. By the time I embarked on the project of writing a master's thesis there were several candidate subjects on my mind; 'meaning' was the one that survived. Although I had never written a decent scientific or philosophical paper before, I decided to produce a comprehensive thesis based on a critical examination of the ideas I had concerning the nature of meaning.

The most important thing that trial and error have taught me in the course of this project is that contributing to science or philosophy has two aspects. The first aspect is having something to contribute. And the second aspect is finding a connection between your contribution and that to which you want it to contribute. The more revolutionary your contribution is, the more difficult – but also the more important – it is to find a connection with existing ideas. The process of finding a connection has been the most difficult and time-consuming part of this project, for in the process of connecting you are forced to both understand existing ideas as well as their different interpretations and look for a usable interpretation for your connection. In addition to this, the connection you succeed in finding often creates new difficulties with that which you wanted to contribute in the first place. Given my desire to contribute to science or philosophy, I found it difficult to find the proper balance between focusing too heavily on my own ideas and concentrating too much on making a connection with existing ideas.

Another important skill I needed mastering at the start of this project was the art of curtailment. Luckily, I was not the only student in that predicament. The Department of Philosophy offered a course, called 'Scriptie schrijven voor CKI', which taught students how to get started with their master's thesis, and one of its subjects was curtailment. Among other things, the course taught me to zoom in, from the general level of my education to the appropriate level of the subject of my thesis: empiricist mentalist semantic theories. In the process of zooming in, the subjects that

were not directly related to my thesis vanished out of the picture. What remained are the ones that you will find discussed in this master's thesis.

A related issue that initially bothered me a lot was which structure to use. Again, the course 'Scriptie schrijven voor CKI' helped me a lot. Before the course I was trying to find a structure in which all subjects I wanted to discuss would fit. Following the course helped me getting unrelated subjects out of the way and left me with a solid and simple structure for my master's thesis.

As is appropriate at this point, I would like to thank Herman Hendriks for trying to understand and taking the time to analyse the content of this master's thesis. He kept on trying to make me connect my ideas with existing philosophical and scientific ideas until I became so frustrated that I actually started doing it, if only to make him happy. But what at that moment seemed like giving in, turned out to be a crucial step in the development of my master's thesis that resulted in far more fruitful sessions between him and me. I would also like to thank Menno Lievers and Henriëtte de Swart for their valuable contributions.

I thought writing this master's thesis would be a straightforward process, but it ended up being a struggle between my desire to really contribute something to philosophy or science and my desire to finish my education within a reasonable time span. I started with the focus on the first desire, but as the weeks, months and even years went by, the second desire steadily became more and more dominant. Despite the ensuing eventual practical boost, the perfectionist inside me is still content and proud with the result, for I feel that I nicely managed to connect my personal ideas with existing ideas in the field.

Contents

CHAPTER 1 INTRODUCTION	5
1.1 <i>UNDERSTANDING UNDERSTANDING</i>	5
1.2 <i>RESEARCH QUESTION</i>	8
1.3 <i>RELEVANCE</i>	9
1.4 <i>THE SEMANTIC PHENOMENON</i>	11
1.5 <i>LOCKE'S EMPIRICIST MENTALIST SEMANTIC THEORY</i>	14
CHAPTER 2 INTERSUBJECTIVITY	21
2.1 <i>THE INTERSUBJECTIVITY OBJECTION</i>	21
2.2 <i>MEETING THE INTERSUBJECTIVITY OBJECTION</i>	30
CHAPTER 3 SELF-APPLICABILITY	46
3.1 <i>THE SELF-APPLICABILITY OBJECTION</i>	46
3.2 <i>MEETING THE SELF-APPLICABILITY OBJECTION</i>	55
CHAPTER 4 ABSTRACT IDEAS	67
4.1 <i>THE ABSTRACT-IDEAS OBJECTION</i>	67
4.2 <i>MEETING THE ABSTRACT-IDEAS OBJECTION</i>	76
CHAPTER 5 CONCLUSION	89
5.1 <i>FINDINGS</i>	89
5.2 <i>FURTHER RESEARCH</i>	92
REFERENCES	96

Chapter 1 Introduction

1.1 Understanding Understanding

The process of understanding – or assigning meaning to – words, sentences and other phrases is not as self-evident as it may seem to competent users of natural language. This is reflected by the fact that the nature of this process has been the topic of extensive scientific debate during the last couple of centuries. We shall use the phrase ‘the semantic phenomenon’ for the assignment of meaning to a sign. The semantic phenomenon is especially visible within natural language, where the words, sentences and other phrases constitute the signs, while their meaning is the information they convey. However, it is also less obviously observable in situations involving traffic signs, disease symptoms and weather forecasts.

The interdisciplinary field of Cognitive Artificial Intelligence aims at making artefacts intelligent by imitating human cognitive capacities. It tries to do so by constructing models of these capacities and implementing them in artefacts (often computers). Artefacts are already capable of successfully displaying several cognitive capacities. Artificially intelligent computers that recognize and respond to spoken language replace human telephone operators; artificially intelligent computers that recognize violent behaviour by audiovisual means assist patrolling policemen; and artificially intelligent computers that play chess beat human Grandmasters. In spite of this, however, the process of understanding has not yet been successfully mimicked in artefacts, despite the importance that it might have for various research areas within the field of Cognitive Artificial Intelligence. Thus the artificially intelligent computers that recognize and respond to spoken language still lack any understanding of what they are recognizing and responding to; and the chess computers have no clue what they are reasoning about. Finding an adequate model of the human capacity of assigning meaning, then, might boost various research areas and probably take Cognitive Artificial Intelligence a step further towards the realization of intelligent artefacts.

A common way of finding a model within Cognitive Artificial Intelligence proceeds by drawing from one or more of its sub-disciplines. A good starting point for looking for a model of the semantic phenomenon may be found in the philosophical

debates about meaning, which date back as far as the ancient Greek. This long history might supply satisfactory models that account for with a lot of empirical data concerning the semantic phenomenon. Philosophy is also known for the strictness and rigour of its analyses, properties that might facilitate the transition to an eventual computer implementation of a model.

There are several popular modern semantic theories within philosophy that might be fruitful for Cognitive Artificial Intelligence, such as the ones proposed by Gottlob Frege and Ludwig Wittgenstein in the latter's *Philosophical Investigations*. Now, most of such modern semantic theories are explicitly non-mentalist. Non-mentalist theories exclude mental properties from their models of the semantic phenomenon. This exclusion is motivated by the claim that such mental properties, although their existence is not denied, are idle wheels that have no role to play in a philosophical analysis of the semantic phenomenon. This is also made clear by Bennett: "They [inner accompaniments] may have a role in the scientific explanation or the natural history of meaning phenomena, but they do not figure in the analysis of the concept of meaning" (Bennett 1971, p. 7.)

Ignoring mental properties, however, might raise problems for the conversion of such non-mentalist theories into models that are compatible with the goals of Cognitive Artificial Intelligence. Whenever a theory of some human cognitive capacity is adopted for implementation on a computer, it is the computer that takes the role of the human subject in that theory. This requires that the mental properties involved in the relevant cognitive capacity be actually included in the theory, because otherwise it remains unclear how the software within the computer must be modelled in a correct imitation of the cognitive capacity at issue. Since modern philosophical analyses of the semantic phenomenon exclude mental properties, such semantic theories are unsuitable for adoption by Cognitive Artificial Intelligence.

Does this mean that philosophy is unable to contribute to Cognitive Artificial Intelligence with respect to understanding, the specific human cognitive capacity of dealing with the semantic phenomenon? No. The philosophical debate about meaning has a long history and whereas modern non-mentalist semantic theories may seem unsuitable, it is worth while exploring older semantic theories. For it may be noted that before Frege and Wittgenstein introduced their non-mentalist approaches to semantics, mentalist semantic theories were the predominant ones. As opposed to

their non-mentalist opponents, such theories did include mental properties into their theories of the semantic phenomenon. Famous mentalist semantic theories have been put forward by René Descartes, John Locke, George Berkeley, and John Stuart Mill.

Traditionally the domain of mentalist semantic theories is partitioned into a rationalist and an empiricist sub-domain. Rationalists such as Descartes believe that the mental is furnished with innate ideas, so that subjects obtain new knowledge by applying the furniture of their mind to reality; while empiricists believe that the mental is a *tabula rasa*, which means that all new knowledge must come from outside via experience. Locke's semantic theory is an empiricist one. It was introduced in the third book of *An Essay Concerning Human Understanding*, which was originally published in 1690. Whether Locke intended to give a semantic theory or rather a theory of knowledge is not entirely clear, but his work includes at least some proposals that might be characterized as components of an empiricist mentalist semantic theory. Just as other contemporary philosophers, Locke proposed ideas as a central ingredient of his theory. Ideas, which had been popularized by René Descartes, were quite vague entities residing in the human mind. According to Locke, such ideas are intermediaries between words and things in reality.

Since it includes mental properties, Locke's empiricist mentalist semantic theory seems a good starting point for finding a model of the semantic phenomenon that can be used in Cognitive Artificial Intelligence. However, it was not without reason that Locke's semantic theory, which had been quite popular for a long time, was abandoned for non-mentalist semantic theories such as Frege's and Wittgenstein's. For Frege and Wittgenstein did not only themselves introduce semantic theories, but also formulated serious objections to Locke's empiricist mentalist semantic theory. The three most important objections can be characterized as the intersubjectivity objection, the self-applicability objection and the abstract-ideas objection. Locke's empiricist mentalist semantic theory cannot be reliably adopted as a Cognitive Artificial Intelligence model of the semantic phenomenon until these three fundamental objections have been met. Only then can it be checked for compatibility with the goals of Cognitive Artificial Intelligence.

The intersubjectivity objection argues that empiricist mentalist semantic theory cannot explain how two subjects understand each other when both have distinct ideas. This objection can be met, according to Jonathan Lowe, by finding an

“intersubjective, publicly available criterion for the similarity of ideas occurring in two different minds” (Lowe 1995, p. 151). What will be argued in Chapter 2 is that what we will call web-similarity, a notion based upon the concept of a web-of-ideas, might supply the intersubjective, publicly available criterion for similarity that Lowe is hinting at.

The self-applicability objection argues that ideas within subjects somehow have to pop up spontaneously if the relevant situations arise. In other words: ideas have to be self-applicable and this is, so the objection goes, highly implausible. In Chapter 3 it will be defended, again with the help of Jonathan Lowe, that ideas are indeed self-applicable. And if the application criteria of ideas are somehow similar to the application criteria of coin-slots in an automatic vending-machine, this might not be so implausible after all.

The last objection involves abstraction and argues that empiricist mentalist semantic theory has problems explaining ideas of abstract entities such as the concept of a triangle. Chapter 4 will try to show that conceiving of ideas as pattern-like entities rather than images, mental objects or stereotypes, may contribute to an adequate account of ideas of abstract entities.

1.2 Research Question

The present master’s thesis sets out to answer the following question:

Is it possible to meet the intersubjectivity objection, the self-applicability objection and the abstract-ideas objection against empiricist mentalist semantic theory within an interdisciplinary approach, and if so: how?

This main research question will be split up into the following three sub-questions, which will be addressed consecutively:

- (1) Is it possible to meet the intersubjectivity objection against empiricist mentalist semantic theory within an interdisciplinary approach, and if so: how?
- (2) Is it possible to meet the self-applicability objection against empiricist mentalist semantic theory within an interdisciplinary approach, and if so: how?

- (3) Is it possible to meet the abstract-ideas objection against empiricist mentalist semantic theory within an interdisciplinary approach, and if so: how?

This thesis can be seen as a first step of a two-step process. This first step is an attempt at finding a potentially suitable candidate model of the semantic phenomenon. The second step would be checking whether such a model is compatible with the goals of Cognitive Artificial Intelligence. Even if the first step (and thus this master's thesis) succeeds and there will turn out to be theories that satisfactorily meet the three main objections against an empiricist mentalist semantic approach, it remains unclear whether step two will succeed in that such theories are compatible with the goals of Cognitive Artificial Intelligence: answering the latter question is beyond the scope of the present thesis.

Before embarking on answering its main research question, this thesis will shortly introduce the semantic phenomenon in general and discuss Locke's empiricist mentalist semantic theory in detail. But before that, we will first discuss the relevance of the main research question.

1.3 Relevance

The present section deals with the relevance of the main research question. It will explain why the question is worth asking and why it has not been satisfactorily answered yet. Observe that the issue is how three objections against a theory about a human cognitive capacity can be met when they are approached from the interdisciplinary perspective of Cognitive Artificial Intelligence. The value of having an answer to this question depends on the components from which it is built up.

About the interest of the object of investigation, empiricist mentalist semantic theory, several things can be said. A first reason why semantic theory in general is interesting is that other disciplines are dependent on it. Computational linguistics is one of these disciplines. (Other fields of study that depend on, or have close connections to semantic theory are, among others, the theories of knowledge and reasoning.) Several features of language have already quite successfully been implemented on computers, but meaning, the connecting phenomenon between speaking and hearing and between writing and reading is still an elusive phenomenon. Scientific progress in semantic theory might accelerate the development towards computers that actually *understand* language. Computers are able to process spoken

sentences and parse them into electronic texts. They can even pronounce sentences in such a way that humans can understand them. This is already a great achievement, but the computers executing these tasks have no sense of what they are processing or producing; they completely fail to grasp its meaning. One of the reasons for the elusive character of meaning might be that there have not been serious attempts to achieve an artificial variant of the cognitive capacity of understanding.

While any investigation into semantic theory in general is interesting on account of the reasons just mentioned, it can be argued that research directed at its *empiricist mentalist* version is of special interest. First, as noted earlier, the *mentalist's* actual inclusion of mental properties in the theory may simplify its transformation into a model that is compatible with the goals of Cognitive Artificial Intelligence: the computer will take the role of the human subject in the implementation of the model, and inserting mental properties in its software may enable it to correctly imitate the human cognitive capacity of understanding. And second, the *empiricist's* interpretation of the semantic phenomenon is relatively computer-friendly in comparison to other approaches. The *tabula rasa* view of sensation lends itself to a straightforward implementation in the input/output architecture of computers, so that a computer may be very fit for the subjects' role within an empiricist setting. It could be made the carrier of signs, ideas as well as the relations between them, and the possession of ideas could make it sensitive to meaning.

The three objections raised against empiricist mentalist semantic theory are interesting because they are obstacles that keep on preventing the theory from being re-accepted. Meeting these objections would remove three major obstacles to acceptance of this particular interpretation of the semantic phenomenon as a plausible one that can be used in Cognitive Artificial Intelligence. The interdisciplinary approach of Cognitive Artificial Intelligence is interesting because it combines the old wisdom and experience of philosophy with modern, fresh knowledge from disciplines such as psychology, computer science, linguistics and neuroscience. All three aforementioned objections have their origin in philosophy from before 1950. This means that they originate from an era when computers and internet were not as integrated into society as they are now; when there was little known about the human brain and its electro-chemical properties; and when artificial intelligence had not yet shown the potential of algorithms in the simulation of certain cognitive capacities. A specific advantage of

the computer era is that computers may make extremely strong demands on the models that are implemented in them. Computer scientists have been struggling with these demands for more than half a century now and have accumulated a lot of information, analyses, experience and tools that may help meet those demands. Using this information, analyses, experience and tools might contribute to the computer-compatibility of empiricist mentalist semantic theory.

From the previous section it should be clear that no material progress has been made since Locke introduced his empiricist mentalist semantic theory; people still refer to his work when talking about empiricist mentalist semantic theory and the objections against it surfaced almost a century ago with Gottlob Frege and Ludwig Wittgenstein, or even longer ago with George Berkeley. The theory and the objections belong to another era, when computers were unknown and the human brain was even more of a mystery than it is now. Applying a modern, interdisciplinary approach could yield an updated version of the empiricist mentalist semantic theory that is immune to the objections that have haunted its predecessor. If this will appear to be impossible in view of the fact that the objections against the theory turn out to be insuperable, we will at least be able to say that empiricist mentalist semantic theory did not give up without a fight.

1.4 The semantic phenomenon

This section shortly explains what will be meant by ‘the semantic phenomenon’ and by ‘information sharing’ in the present thesis.

Imagine a situation where a streetlight flashes when someone is passing by. Imagine, moreover, that this person has recently lost his lover. In this situation, the flashing of the streetlight may mean to the passing person that his lost lover tries to contact him. Now imagine the situation where there are red spots all over your body. These red spots may mean that you have the measles. Both situations demonstrate what we will call the semantic phenomenon. In both situations there is a sign: the flashing streetlight and the red spots, respectively, that stands for certain information: a deceased lover who is trying to contact the passing person or the fact that you have the measles. This is the phenomenon that a semantic theory should clarify: a sign that stands for certain information (see also Figure 1).

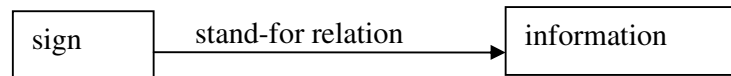


Figure 1

So the key ingredients of the semantic phenomenon are a sign, the stand-for relation and certain information. Signs may have many different forms, and may include flashing streetlights and red spots, but the most obvious example of signs are the words of natural language. Thus common nouns such as ‘snow’, ‘tree’ and ‘democracy’ are all signs that stand for certain information. Such words will also constitute the paradigmatic examples of signs studied in this thesis. For the purpose of expository simplicity, other types of linguistic signs will be left outside of consideration. This means that, among many other interesting types of signs, for example the articles ‘the’ or ‘a(n)’ and indexical signs such as ‘I’, ‘tomorrow’ and ‘here’ will be ignored throughout this thesis. Another assumption that should be mentioned is that words, and not sentences, are taken to be the fundamental units of meaning. This means that sentences have a meaning because they are built up from meaningful words and that the meaning of a sentence is a function of, among other things, the meanings of the words of which it consists.

As explained above, when a theory of a human cognitive capacity such as understanding is adapted for implementation on a computer, it is the computer that takes the role of the human subject in that theory. And this requires that the mental properties involved in the semantic phenomenon be actually included in the theory, because otherwise it remains unclear which properties the computer software should exhibit. Because of this, we will not pay attention to popular non-mentalist semantic theories in this thesis. Instead, we will follow mentalist semantic theories that were popular between the 17th and 19th century and this already enables us to fill in some blind spots of the semantic phenomenon. As regards the nature of the information for which a sign stands in the semantic phenomenon, we may conclude that information is a mental entity within the subject’s mind. Although this is still quite vague, it rules out any interpretation that locates the information somewhere outside the subject. A

more exact interpretation will be developed in the process of answering the main question of this thesis.

The third important ingredient of the semantic phenomenon is the stand-for relation that signs have with the information. Explaining the nature of this relation is explaining how a sign activates that information, in other words, the way in which a sign determines the information that belongs to it. Our adoption of a mentalist semantic framework again restricts the possibilities: the information determined by the stand-for relation of some sign must be some mental entity. Since information is a mental entity, mentalist semantic theories are forced to include an additional relation into their accounts: one between mental entities and the external world. Roughly, this relation ensures that the mental entities are correctly applied to objects and other phenomena in the external world.

The semantic phenomenon plays an indispensable role in what is known as information sharing. Information sharing is a common concept in science, where it appears in many forms. It figures in debates within philosophy, linguistics, computer science and biology. The concept involves (at least) two subjects that share information about the world (see also the picture below). Our employment of the phrase ‘subject’ instead of ‘human’ serves to highlight the fact that information sharing is not restricted to humans. Most animals have ways of sharing information, albeit in a more primitive form than humans, and also computers are capable of information sharing. Two subjects that share information both use some agreed sign. Note that this agreement should be taken in a very broad sense, including unconscious agreements that are dictated by evolution. Examples of information sharing are not hard to find. Two dogs may share information using the position or the movement of their tail as the agreed sign. Two computer applications may share information using bits and bytes as their agreed sign. In the latter case the agreement about which sign stands for which information is called a protocol. A very common computer protocol is HTTP (Hyper Text Transfer Protocol), which is used for computers that share Internet sites.

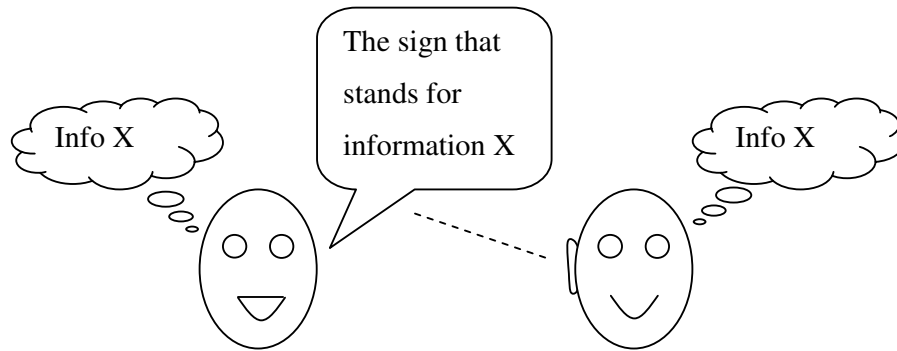


Figure 2

The semantic phenomenon comes down to a subject having a specific sign stand for certain information. Information sharing actually involves two occurrences of that phenomenon. It is the special case of the co-occurrence of two semantic phenomena in different subjects where the two subjects have one and the same sign stand for the same information. Whenever two subjects agree to choose the same sign as stand-in for the same information, the conditions of information sharing are met. The semantic phenomenon is thus a necessary, but not a sufficient, condition for information sharing, which means that situations in which the semantic phenomenon occurs may fail to be situations of information sharing.

1.5 Locke's Empiricist Mentalist Semantic Theory

The present section will give a detailed description of empiricist mentalist semantic theory with the focus on the one proposed by John Locke. It will include a short historical introduction, the empiricist mentalist interpretation of the semantic phenomenon, as well as the most important objections raised against empiricist mentalist semantic theories.

As already mentioned, empiricist mentalist semantic theories came in the wake of theories of knowledge that were proposed by famous philosophers such as George Berkeley, René Descartes, David Hume and John Locke in the 17th and 18th century. Empiricists such as John Locke tried to base their theory of knowledge on the experience of a subject as opposed to, for example, René Descartes, who founded the knowledge of a subject on its ratio. In his famous *An Essay Concerning Human Understanding* (Locke 1690) John Locke derives his empiricist mentalist semantic theory from a theory of knowledge built around ideas that furnish the human mind.

Locke's book received considerable positive and negative criticisms from his contemporaries and is still considered a philosophical classic. This master's thesis will treat Locke's book as a representative example of empiricist mentalist semantic theories in general¹ and all proposed adaptations inspired by various objections are supposed to apply to John Locke's theory. As indicated above, in the 20th century this theory, which until then had been very popular, became outdated when a new era started with the development of alternative semantic theories by philosophers such as Gottlob Frege and Ludwig Wittgenstein.

The purpose of this master's thesis requires that the semantic phenomenon that was introduced above be described in further detail. This description will have to include the interpretations that are specific to John Locke, our representative of empiricist mentalist semantic theory. Below, each of the notions that jointly make up the semantic phenomenon will be treated separately, viz.: respectively, 'sign', 'information' and 'stand-for relation'.

Let us first consider the notion of a sign. This ingredient of the semantic phenomenon is not treated very extensively within the semantic literature. Thus, it is unclear whether empiricists and, more specifically, John Locke interpret it as a phenomenon in reality or as a sensation that derives from some phenomenon in reality. O'Connor hints at this when he says that '[u]nfortunately Locke gives no clear account of what he conceives to be the nature of a sign' (O'Connor 1967, p. 33). Because the precise nature of signs will not play a central role in the present thesis, the interpretation of the concept will not be clarified further. With respect to John Locke's theory it may be noted that his usage of the word 'sign' differs slightly from the way in which we employ it in the context of the semantic phenomenon. According to John Locke both words and ideas are signs (Locke 1690, Section IV.xxi.4): the words are signs of ideas and the ideas are signs of things in reality. In this thesis, however, the word 'sign' only refers to the thing (i.e. word or traffic sign) that is said

¹ Because *An Essay Concerning Human Understanding* by John Locke is used as paradigmatic for empiricist mentalist semantics, this master's thesis will frequently cite from it. These passages will be referred to with three numbers that have the following meaning. In 'Section III.ii.1' the first number 'III' refers to the third of the four books of which Locke's Essay consists; the second number 'ii' refers to the second chapter within that book; and the third number '1' refers to the first section within that chapter.

to stand for information, the latter being the counterpart of what are called ideas in mentalist semantic theories. This means that words shall be referred to as signs of ideas, but ideas shall not be referred to as signs of things in reality.

The second ingredient of the semantic phenomenon is information. According to empiricist mentalist semantic theory the information for which a sign stands is a subjective idea. The unclear nature of the latter notion forces the theory to say more about the concept of subjective ideas. The most pressing questions concern their origin and nature, and the most important contribution to their analysis has been made by John Locke. On Locke's account, subjective ideas are spatio-temporally abstracted impressions that reach us via our senses. He specifies several categories of ideas which are to reflect the several types of words (signs) and facts of which the language and knowledge of human subjects consists. Locke's first distinction is that between simple and complex ideas. Simple ideas cannot be analysed further into other ideas and thus are non-compounded (Locke 1690, Section II.ii.1). Complex ideas, on the other hand, are compounded and thus analysable into other ideas. Complex ideas are the result of activities of the subject's mind aimed at combining ideas into new ones. John Locke distinguishes three different types of complex ideas: modes, substances and relations. Modes are 'complex ideas which, however compounded, contain not in them the supposition of subsisting by themselves but are considered as dependences on, or affections of substances' (Locke 1690, Section II.xii.4). Substances are 'such combinations of simple ideas as are taken to represent distinct *particular* things, subsisting by themselves' (Locke 1690, Section II.xii.6) and a relation 'consists in the consideration and comparing one idea with another' (Locke 1690, Section II.xii.7). Locke further sub-divides modes into simple and mixed ones.

An important difference between mentalist and non-mentalist semantic theories is the mentalists' postulation of subjects as a necessary condition for the semantic phenomenon to arise. Several additional questions arise with this postulation of subjects as the carrier of ideas. Note that in the case of languages used by human subjects, the words are thought to be somehow related to reality; this is what is also known as the 'aboutness' of natural language. Some non-mentalist semantic theories justify this intuition by interpreting the information expressed by words as something that resides in reality, but because of its postulation of subjects as the carriers of ideas – i.e., information – mentalist semantic theory is obliged to provide an alternative

justification of the ‘aboutness’ intuition. Accordingly, mentalism says that subjective ideas relate to reality. Because ideas reside inside a subject and because subjects, at least human subjects, do not possess these ideas from birth, the ideas must somehow develop inside the subject. O’Connor refers to this as the process of manufacture (O’Connor 1967, pp. 42-43). Instead, we will refer to this process as the *development process of ideas*. This is done in order to emphasize its (possibly) unconscious character, so that readers will be prevented from interpreting it as a process of conscious creation with all sorts of concomitant problems. While the development process of ideas explains how ideas are generated by experience of the subject, another question is how these ideas are used by the subject after the development process has ‘finished’. This second question refers to a process in which a certain situation or object in the material world triggers or activates an already existing idea; this process will be called the *activation process of ideas*. According to Locke, a situation or object in reality makes an idea active whenever it ‘agrees’ or is ‘observably similar’ to that idea (Locke 1690, Section III.iii.13). Unfortunately, he does not elaborate on what this ‘agreement’ may involve. These very important notions can be elucidated by considering a subject without any ideas. The development process of ideas explains how ideas come to exist within this subject. John Locke mentions several activities of the mind that develop ideas of the categories of modes, substances and relations mentioned above, such as abstracting, combining and sorting. After some idea has been developed it is ready for the process of activation. The latter process takes care of the fact that the ideas developed within the subject are activated by those situations in reality by which they have come to be triggered during the development process. According to Locke there are two sources from which the development process extracts ideas: sensation and reflection. The first source concerns the signals received from the sense organs of the subject, such as the visual sense organ. The second source of ideas concerns ‘*the internal Operations of our Minds*’ (Locke 1690, Section II.i.2) and should, according to some (such as Rabb 1985, p. 14), not be interpreted as a conscious internal sense organ, but rather as a reflex based on the internal operations.

Before discussing the third ingredient of the semantic phenomenon, three more issues need to be addressed. First, note that the above description of the activation and development process of ideas may suggest that they are clearly distinguished, for

example in a temporal sense. Thus the development process takes place first and, after this process is finished, the activation process can start. Although this order may actually occur, it is not necessary. For example in the case of human subjects, the distinction might not be that clear. It could be that the activation and development process of ideas actually run in parallel. And how to realize both processes in, for example, a computer is something that must be decided by the designer of the software it runs on. The second issue concerns an essential consequence of the mentalist's interpretation of the information involved in the semantic phenomenon. This consequence is that whenever in some situation the meaning of a sign is explained by a mentalist, this should always be done relative to some subject. Hence questions such as "What is the meaning of the word 'tree'?" should, in the context of mentalist semantics, be further specified as "What is the meaning of the word 'tree' *for this or that subject*?" because the existence of the relation between some word and its meaning depends on the subject that carries it. Ignoring this could cause unnecessary problems. Finally, the third issue involves the term 'idea' that has been used to refer to the empiricist mentalist interpretation of the concept of information within the semantic phenomenon. In order to enhance readability, we will continue to employ this term in the remainder of this master's thesis for referring to the specific interpretation that empiricist mentalist semantics assigns for the concept of information. On the other hand, the term 'information' will denote the corresponding concept as it occurs in the semantic phenomenon in general.

The third and last ingredient of the semantic phenomenon is the stand-for relation. We have already discussed the way in which empiricist mentalist semantic theory interprets signs and information. In addition, the stand-for relation is supposed to capture the way in which the concepts of signs and information are related and indicate how given a sign, the corresponding information is chosen or triggered. According to empiricist mentalist semantics this relation is conventional in the sense that the stand-for relation between signs and (in this case) ideas is a convention or agreement between subjects (Locke 1690, Section III.ii.1). This contrasts with so-called natural relations, such as the one between fire and smoke, that do not rely upon agreement between subjects, but upon causal laws of nature. It may be noted that John Locke does indeed go into the relation between signs and information, for instance when he says that "[t]he use [...] of Words, is to be sensible Marks of *Ideas*; and the

Ideas they stand for, are their proper and immediate Signification’ (Locke 1690, Section III.ii.1). Although Locke recognizes this relation, he does not go into the nature of what we will call the stand-for relation between signs and ideas. Note that this relation must be distinguished from the relation between ideas and reality discussed above, and hence also from the development and activation process of ideas, since the latter concepts mediate between reality and ideas, whereas the stand-for relation holds between signs and ideas.

As already explained, an actual implementation of Locke’s empiricist mentalist semantic theory as a model that is to be used in Cognitive Artificial Intelligence is beyond the scope of the present master’s thesis. Before it can be used, moreover, the most important objections raised against it should be met. This thesis will investigate the possible ways in which this might be done. The objections are three in number, and they concern the empiricist mentalist interpretations of the three main ingredients, sign, information and stand-for relation, of the semantic phenomenon.

The first objection against the empiricist mentalist semantic theory is that it cannot explain intersubjectivity. Intersubjectivity is the phenomenon that two subjects ‘understand’ each other, because they both have one and the same sign stand for the same information. This is a problem for mentalist semantics because the theory explains the information for which a sign stands in terms of subjective ideas. This means that subjects do not ‘understand’ each other in the straightforward sense that they both have their sign stand for the *same* information. Thus there cannot be intersubjectivity. Because intersubjectivity is a common phenomenon in human language it should be accounted for. Therefore, if empiricist mentalist semantic theory fails to explain it, the theory cannot be correct. This first objection will be called the *intersubjectivity objection*.

The second objection against empiricist mentalist semantic theory focuses on the activation process of ideas. This process involves the activation of an idea by some situation in the material world. According to Locke (1690, Section 3.3.13) this activation is based upon similarity, so the idea is observably similar to the situation in the material world by which it is activated. This image-like conception of ideas requires ideas to be interpreted themselves and this gives rise to an infinite regress of interpreting the interpretation of an interpretation. The only route out of this infinite regress is by attributing a self-applicable character to ideas which causes them to pop

up automatically without interpretation whenever a situation in the material world arises to which they are observably similar. This route, however, is claimed to be implausible for two reasons. The first reason is that image-like ideas, just like images, cannot be self-applicable and always require interpretation. The second reason is that ideas' supposed self-applicable character can easily be attributed to signs instead without loss of explanatory power and this renders ideas useless. Since empiricist mentalist semantic theory cannot plausibly explain the 'aboutness' of language, it must be incorrect. This second objection will be called the *self-applicability objection*.

The third objection focuses on the development process of ideas. The objection is that this development process of ideas cannot account for certain types of information that signs are known to stand for. If we try to apply the development process to abstract information such as the ideas of 'democracy' or 'triangle' it is not clear how such information may be abstracted on the basis of sensation. The objection is that since such information cannot be abstracted from sensation and since it is information for which certain signs stand – namely the words *democracy* and *triangle* – empiricist mentalist semantics is wrong. This third objection will be called the *abstract-ideas* objection.

Each of the next three chapters will focus on one of these objections and investigate whether and, if so, how it might be met by empiricist mentalist semantic theory.

Chapter 2 Intersubjectivity

This chapter focuses on the first sub-question of the present master's thesis, viz., is it possible to meet the intersubjectivity objection against empiricist mentalist semantic theory within an interdisciplinary approach, and if so: how? The chapter consists of two sections. The first one gives a detailed description of the intersubjectivity objection against empiricist mentalist semantic theory. Two versions of this objection will be treated and confronted with Locke's thoughts about intersubjectivity. Correspondingly, it will be argued that two problems lie at the heart of the intersubjectivity objection and some existing proposals for solving these problems will be discussed and evaluated. The second section addresses the two central problems in the context of the existing proposals for solving them, proposes some further solutions, and investigates the consequences of these solutions for empiricist mentalist semantic theory in general.

2.1 The Intersubjectivity Objection

In the previous chapter we shortly mentioned the objection that empiricist mentalist semantic theory cannot account for intersubjectivity. The present section consists of five parts. The first part focuses on intersubjectivity, explains why it is such an important notion and outlines how it relates to semantic theory in general. The second part discusses further the objection against empiricist mentalist semantics that it lacks intersubjectivity. In this second part we will review Wittgenstein's and Frege's arguments for the claim that intersubjectivity is a problem for empiricist mentalist semantic theory. The third part shows that the objection does indeed apply to the work of the prominent empiricist mentalist John Locke. The fourth part discusses two proposals in the literature that try to meet the intersubjectivity objection. Finally, the fifth part critically analyses these proposals.

2.1.1 Intersubjectivity

Intersubjectivity, explained from the perspective of the information-sharing framework described above, arises if and only if both subject A and subject B have a sign *S* stand for the same information *I*. This means, for example, that John and Mary reach intersubjectivity about the word 'tree' if and only if the word 'tree' stands for

the same information for both of them. This informal explanation shows why intersubjectivity is such an important notion. Imagine that intersubjectivity were to fail with respect to the word 'tree' as used by John and Mary. This would mean that if John were to use the word 'tree' when talking to Mary (or vice versa), they would not understand each other because they would both associate different information with the word 'tree'. Thus, without intersubjectivity communication is impossible.

In a previous section, semantic theory was required to interpret the most important concepts of the semantic phenomenon: sign, information and the stand-for relation. That section also pointed out the relation between information sharing and the semantic phenomenon and explained that situations of information sharing always involve the semantic phenomenon, but that situations where the semantic phenomenon occurs do not always involve information sharing. In other words, the semantic phenomenon is a necessary but not a sufficient condition for information sharing, just like clouds are a necessary but not a sufficient condition for rain, the latter requiring that further conditions pertaining to temperature and humidity be satisfied. Thus, in a situation where two subjects A and B have one and the same sign S stand for certain information (so that the semantic phenomenon occurs twice) there is an additional condition, viz., the condition that both A and B have S stand for one and the same bit of information (they have to 'agree' about S), to be met in order for A and B to actually partake in information sharing.

2.1.2 The Objection and Two Derived Problems

Opponents of empiricist mentalist semantic theory observe that intersubjectivity is very common in natural language. Given this observation, they argue that empiricist mentalist semantic theory must be wrong since it fails to account for the important and common phenomenon of intersubjectivity.

The intersubjectivity condition, as illustrated above, seems to require that the information for which a sign stands should be somehow publicly available. Ludwig Wittgenstein claimed that this requirement is not met by mentalist semantic theory. According to Candlish (2004, Section 'The Significance of the Issue'), "the implication [made by empiricist mentalist semanticists, among others] is that the internal vehicle of my musings could in principle be private". Intersubjectivity (or 'agreement') plays a very fundamental role in Wittgenstein's interpretation of the

semantic phenomenon, because his account claims that the *publicly accessible use* of a sign determines for which information it stands. This is also stressed in the following quote from Candlish (2004, Section ‘The Significance of the Issue’):

Wittgenstein suggests that the existence of the rules governing the use of language and making communication possible depends on agreement in human behaviour – such as the uniformity in normal human reaction which makes it possible to train most children to look at something by pointing at it. [...] One function of the private language argument is to show that not only actual languages but the very possibility of language and concept formation depends on the possibility of such agreement.

So according to Wittgenstein the semantic phenomenon is impossible without intersubjectivity (because intersubjectivity comes in at the level of information), and in order to show this he presents his famous private language argument (Wittgenstein 1953, Sections 243-271). The exact implications of this argument are under discussion and the argument itself has apparently been misinterpreted more than once,¹ but these issues will be ignored here since the question of the ‘correct’ interpretation is beyond the scope of this master’s thesis. Instead, we will simply build on the interpretation supplied by Candlish.

Note the difference between Wittgenstein’s intersubjectivity as a *necessary condition* for the semantic phenomenon and the above view of intersubjectivity as an *additional condition* to it. So where Wittgenstein interpretation of the semantic phenomenon states that meaningfulness and intersubjectivity go hand in hand, our interpretation gives meaningfulness (the semantic phenomenon) a more fundamental role than intersubjectivity. This difference has consequences for the intersubjectivity objection. The conclusion of Wittgenstein’s objection as Candlish (2004, Section ‘Overview: Wittgenstein’s Argument and its Interpretations’) explains it is “that a language in principle unintelligible to anyone but its originating user is impossible. The reason for this is that such a so-called language would, necessarily, be unintelligible to its supposed originator too, for he would be unable to establish meanings for its putative signs.” So, whereas the standard objection argues that mentalist semantic theory fails to account for intersubjectivity and is therefore unable to explain how two subjects communicate, Wittgenstein’s private language argument

¹ See Cook (1965).

states that, because of this fundamental role for intersubjectivity, the private character of the information in mentalist semantic theory renders not only communication between *two* subjects impossible, but excludes the possibility of communication within *one* subject as well. In other words, the private language argument states that purely private semantic theories, in which the information for which a sign stands is *necessarily* inaccessible to other subjects, are inadequate because they make the semantic phenomenon inaccessible (or incomprehensible or unintelligible or unknowable), not only to others but also to the subject itself!

Another important opponent of mentalist semantic theory who used intersubjectivity to show its deficit is Gottlob Frege. Picardi (1996, Section ‘Frege’s Anti-Psychologism’) quotes the following passage from a draft of one of his letters:¹

Now if the sense of a name was something subjective then the sense of the proposition in which the name occurs and hence the thought would also be something subjective and the thought one man connects with this sentence would be different from the thought another man connects with it; a common store of thoughts, a common science would be impossible. It would be impossible for something one man said to contradict what another man said, because the two would not express the same thought, but each his own. For these reasons I believe that the sense of a name is not something subjective [~~crossed out: in one’s mental life~~], that it does not therefore belong to psychology and that it is indispensable.

Here, the sentence “It would be impossible for something one man said to contradict what another man said, because the two would not express the same thought, but each his own” most explicitly expresses the intersubjectivity objection against mentalist semantic theory. It is also mentioned in Jolley (1999, pp. 163-164). Two variants of this objection are described by Michael Ayers (1991, pp. 277-279).

We may describe the objection in terms of the semantic phenomenon and information sharing: an idea is subjective, hence, as mentalism has it, if for a subject A, sign S stands for (A’s) idea I_1 while for another subject B, the sign S stands for (B’s) idea I_2 , then the two subjects are involved with different ideas: $I_1 \neq I_2$. The sign

¹ Although this citation only considers ‘the sense of a name’, similar considerations hold for senses of other linguistic entities such as predicates.

S therefore necessarily stands for different bits of information in both cases.¹ The intersubjectivity condition described above requires that for two subjects, the sign used by them stands for the same information. Since this is obviously not the case, mentalist semantic theory cannot meet the condition of intersubjectivity and therefore fails to explain how language serves its purpose of information sharing. And if it cannot explain how language works, it is an inadequate semantic theory.

The intersubjectivity objection poses two problems for proponents of mentalist semantic theory. The first is the problem whether and how two subjective ideas can ever satisfy the condition of intersubjectivity, since this implies that they should somehow instantiate the same information. And this brings us to the question of how ideas and information are related? Until this question is answered, it remains unclear how the intersubjectivity condition could possibly be met within a mentalist semantic framework. The second problem is how proposals for solving the first problem can be realized in practice. A solution to the first problem involves only the possibility of a situation in which the intersubjectivity condition is met, but it does not explain *whether* and *how* this situation can actually arise. Solving the second problem, then, requires an explanation of how two subjects may reach a state where they have a sign stand for the same information. A complication for such an explanation is pointed out by Wittgenstein. According to Wittgenstein it is impossible to solve the second problem without already assuming a language (see Crossley 1996, p. 41). More specifically, it is impossible for a subject to explain to another subject which information a certain sign stands for without using language.² For this purpose, Wittgenstein introduces rules which are supposed to be inherent to our ‘form of life’. The following passage from Wittgenstein’s *Philosophical Investigations* illustrates his position (1953, Section 28):

The definition of the number two, “That is called ‘two’” pointing to two nuts, is perfectly exact. But how can two be defined like that? The person one gives the definition to doesn’t know what one wants to call “two”; he will suppose that “two” is the name given to *this* group of nuts! He *may* suppose this; but

¹ As explained above, the subjective ideas are the empiricist’s implementation of what is referred to as information in our description of the semantic phenomenon.

² Here ‘language’ is meant to include ‘ostensive definition’ (definition by pointing) and other explanatory mechanisms.

perhaps he does not. He might make the opposite mistake; when I want to assign a name to this group of nuts, he might understand it as a numeral.

2.1.3 The Objection Applied to Locke

The intersubjectivity objection uses the central importance of communication for language to argue that mentalist semantic theory cannot be correct because it does not give rise to intersubjectivity. In Book Three of John Locke's *An Essay Concerning Human Understanding* the following passages show that the issue of intersubjectivity is indeed not fully explained (Locke 1690, Section III.ii.4):

First, *they suppose their Words to be Marks of the Ideas in the Minds also of other Men with whom they communicate: For else they should talk in vain, and could not be understood, if the Sounds they applied to one Idea, were such, as by the Hearer, were applied to another, which is to speak two Languages. But in this, Men stand not usually to examine, whether the Idea they, and those they discourse with have in their Minds, be the same: But think it enough, that they use the Word, as they imagine, in the common Acceptation of that Language; in which case they suppose that the Idea, they make it a Sign of, is precisely the same, to which the Understanding Men of that Country apply that Name.*

And (Locke 1690, Section III.ii.8):

'Tis true, common use, by a tacit Consent, appropriates certain Sounds to certain Ideas in all Languages, which so far limits the signification of that Sound, that unless a Man applies it to the same Idea, he does not speak properly: And let me add, that unless a Man's Words excite the same Ideas in the Hearer, which he makes them stand for in speaking, he does not speak intelligibly.

While Locke seems to agree with the privacy of ideas (Locke 1690, Section III.ii.1) when he claims that 'they are all within his own breast, invisible and hidden from others, nor can of themselves be made to appear' and talks about ideas within two subjects that are 'the same', he fails to explain how distinct private ideas within different subjects could ever be said to be 'the same' (i.e., the first problem described above).

Wittgenstein argued that it is impossible to explain for which information a certain sign stands without assuming some further language. It cannot be assumed that this explanatory language is innate, however, since this would contradict Locke's opinion of innate ideas (Locke 1690, Section I.ii.1):

It would be sufficient to convince unprejudiced Readers of the falseness of this Supposition [of innate principles], if I should only shew [...] how Men, barely by the Use of their natural Faculties, may attain to all the Knowledge they have, without the help of any innate Impressions; and may arrive at Certainty, without any such Original Notions or Principles.

This passage clearly shows that Locke does not allow for any innate ideas or agreements, which renders his empiricist version of mentalist semantic theory vulnerable to Wittgenstein's attack.

2.1.4 Some Proposals

The reason why intersubjectivity is a problem for empiricist mentalist semantic theory seems to be that intersubjectivity has strong associations with objectivity, whereas empiricist mentalist semantic theory proposes a subjective basis for the semantic phenomenon. And because subjectivity is the opposite of objectivity, opponents of empiricist mentalist semantic theory believe that the two cannot be combined. A reply to this objection is that subjectivity can, under certain circumstances, appear to be objective and that it is this appearance which causes opponents to believe that semantic theory must have an objective basis. These are circumstances where subjects have the same ideas and, moreover, every subject has the relevant signs stand for the same information, so that the intersubjectivity condition is met without there being an objective basis. However, some further problems need to be solved in order for this reply to become plausible. We will now turn to preliminary solutions for the two problems discussed at the end of section 2.1.2.

A proposal for dealing with the first problem can already be found in the literature (Jolley 1999, p. 164; Margolis and Laurence 1999, p. 7). In order to explain whether and how signs can carry the same information if they represent subjective ideas, the type-token distinction is invoked. This distinction is best illustrated with an example. Imagine a child who, as a punishment, writes down one hundred times 'I should not speak when the teacher is speaking' in his exercise book. How many sentences does this child write down? There are two possible answers to this question: 'one' and 'a hundred'. If your answer is 'one', then you interpret the word 'sentence' as a type: the child writes down one type-sentence, but a hundred tokens of it. If your answer is 'a hundred', then you interpret the word 'sentence' as a token. This type-token distinction may also serve to explain how a sign used by two different subjects

may stand for the same information, even if it represents their subjective ideas. On this explanation the ideas for which the sign stand may constitute different tokens that are of the same type – where the latter is the information for which the sign stands. Thus different subjects can have private token-ideas with a shared type.

There have also been proposals for dealing with the second problem (Stokhof 2003, pp. 30-31). While the first problem questions whether it is theoretically possible for two distinct ideas (in different subjects) to carry the same information, the second problem, assuming this theoretical possibility, involves the actual process responsible for realizing this intersubjectivity condition. In natural language, for example, words stand for ideas, and every human has an idea that carries the information TREE. Why is this? Why do all humans have an idea that carries this information? Above we described the development process (as well as the activation process) of ideas.¹ Observe that this development process applies to the second problem, because it somehow ensures that humans develop the same information. A possible answer to the second problem is that the common outside world from which all subjects derive their ideas ensures that these distinct ideas carry the same information: this is due to the fact that they are all about this common outside world. Another component of the answer could be that all (human) subjects extract these distinct ideas by using similar idea-extraction instruments: the human senses and the human brain. So the combination of a common outside world and common idea-extraction instruments may explain how distinct ideas can come to carry the same information.

2.1.5 Do the Proposals Work?

The proposed solution to the first problem exploits the type-token distinction for explaining how two distinct ideas can be said to be similar. Whenever two distinct token-ideas are tokens of the same type-idea they can be said to be similar, just as two distinct token-sentences of the same type-sentence. Although this technically seems to solve the problem, there is the conceptual complication that the nature of the type-ideas remains unclear. Nor is it transparent how they are related to their token-ideas. Observe that explaining the type-ideas in terms of abstract entities would move

¹ Recall from the previous chapter that the development process of ideas might be an unconscious process over which subjects have no direct control.

empiricist semantic theory a lot closer to the semantic theories of Plato and Frege, according to whom such type-ideas live in a third realm: the world of senses, which exists in addition to the objective material world and the world of subjective ideas. This proposal for dealing with the first problem seems a good start, then, but more details are necessary for turning it into a plausible solution.

The proposed solution to the second problem addresses the question how two subjects may reach a state where they have a sign stand for the same information, so that it represents different token-ideas that are actually tokens of the same type-idea. Thus the proposal anticipates a satisfactory solution to the first problem. Above it was suggested that the combination of a common outside world and common idea-extraction instruments may be used to explain that distinct token-ideas in different subjects which have the same sign stand for them are tokens of the same type-idea. Under the assumption that it is in principle possible for two token-ideas to be tokens of the same type-idea this is a promising solution. A possible criticism against this proposal is that it is too weak. Possible opponents could say that the proposed solution only makes it plausible but does not guarantee that the distinct token-ideas are tokens of the same type-idea. Another possible criticism is that the proposal does not fully solve the second problem since it only explains the similarity between token-ideas within different subjects, whereas the second problem also requires an explanation of how (in addition to token-ideas) signs and stand-for relations actually become similar within different subjects. So also for solving the second problem more details are required.

This section has given an overview of the most important features of the intersubjectivity objection against empiricist semantic theory. The term 'intersubjectivity' was explained, the intersubjectivity objection was outlined with the help of Wittgenstein's and Frege's work, and some proposed solutions were discussed. In spite of these proposals, the intersubjectivity objection still puts pressure on empiricist semantic theory. In order to release this pressure, the proposed solutions need to be worked out in more detail. In this master's thesis they will be complemented with new ideas in order to arrive at satisfactory solutions.

2.2 Meeting the Intersubjectivity Objection

This section focuses on the way in which empiricist semantic theory may meet the intersubjectivity objection. A detailed description of this objection was given above, where the objection was partitioned into two main problems that need further clarification. The present section goes further into these problems. The first problem is handled in Section 2.2.1, and Section 2.2.2 discusses how the second problem may be solved.

2.2.1 The First Problem

We have seen that against the background of the semantic phenomenon, intersubjectivity occurs whenever two subjects have the same sign stand for the same information. Now, because empiricist semantic theory interprets this information in terms of subjective entities (viz., ideas), the intersubjectivity condition may seem impossible to reach. If each subject has its own idea, it must necessarily hold that whenever for one subject a sign stands for one idea, then for another subject the same sign inevitably stands for a different idea. On the other hand, for intersubjectivity it is really required that the sign stands for the *same* idea for both subjects. And although Locke assumed this indeed to be the case, this is obviously not possible under the traditional empiricist interpretation of information in terms of subjective ideas. Above it was shown that Frege used this as an argument against mentalist semantic theory. We also suggested a preliminary solution built on the possibility that the two subjective ideas may be different tokens of some common type-idea. It was also noted that this solution only solves part of the problem, because the status of the common type-idea fails to be clarified. The type-idea may be taken to be some abstract entity, but then the account begins to look dangerously much like a Platonist semantic theory. Hence, we concluded that the preliminary solution is in need of further details regarding these type-ideas. Accordingly, the present section tries to meet the intersubjectivity objection by closely examining the definition of the concept of similarity that plays such a crucial role in it. For note that the type-token distinction can also be reduced to a question about similarity, since it is a standard assumption that two different tokens that are of one and the same type are 'essentially the same', or more strictly speaking: similar to one another.

Intersubjectivity requires that signs used by different subjects stand for the *same* information. Within empiricist semantic theory this information finds itself at the level of subjective ideas, which means that intersubjectivity requires that two subjective ideas be the same. At first glance it may seem puzzling how *two* such ideas can ever be ‘the same’. If, for example, this requires that the ideas occupy the same spatio-temporal location, then two subjective ideas at different spatio-temporal locations cannot be similar. But other definitions of similarity are possible, and maybe even more plausible. Jonathan Lowe emphasizes that “in saying that two particulars ‘resemble’ one another, it is always necessary to specify *in what respect* they do so...” (Lowe 1995, p. 163). Others also stress the importance of similarity and the need of a strict definition (Bechtel and Abrahamsen 1991, p. 121; O’Connor 1967, p. 142). Although Lowe acknowledges the importance of similarity, he adds that “it is hard to see how there can be any *intersubjective*, publicly available criterion for the similarity of ideas occurring in two different minds” (Lowe 1995, p. 151). But this is exactly what we will suggest: another definition of similarity, which can be checked on the basis of intersubjective, publicly available criteria. And it will be argued that this notion of similarity, which we will call web-similarity, may be more appropriate in the context of intersubjectivity than the concepts which are usually adopted. Web-similarity will be suggested as a solution to the problem of how the intersubjectivity condition can be achieved by empiricist semantic theory. Using it, the intersubjectivity condition simply comes down to two subjects having a sign stand for web-similar ideas.

Before more specifically going into web-similarity it is necessary to consider webs-of-ideas. Empiricist mentalist semantic theory attributes a collection of private ideas to every subject. These private ideas reside inside the subject and cannot be directly accessed by other subjects. Moreover, each idea can be assumed to have a content that is also inaccessible for other subjects. We suggest that the content of an idea can be identified with its position in the web-of-ideas, which is determined by the way in which the idea is related to other ideas and eventually, via the sensory-motor system, to the outside world. So every subject has a collection of private ideas which owe their content to the way in which they are related to one another and to the outside world; this collection of interrelated private ideas is what we will call a subject’s *web-of-ideas*. On the basis of the above, then, a web-of-ideas can be

formalized as an ordered pair $\langle A, R \rangle$, where A is some set of ideas and R is an irreflexive binary relation on A , that is: $R \subseteq A \times A$ and for all $a \in A$ it holds that $\langle a, a \rangle \notin R$.

Given such structures, the notion of web-similarity intuitively comes down to the following: ideas of possibly different subjects are web-similar if and only if they occupy ‘the same’ position within their own webs-of-ideas. In other words: within their own webs-of-ideas, web-similar ideas have to be related in the same way to all other web-similar ideas.

Formally, this comes down to the following. Let $\langle A, R \rangle$ and $\langle B, S \rangle$ be two webs-of-ideas. Now, since being web-similar means occupying ‘the same’ position within your own webs-of-ideas, it seems reasonable to assume that every idea in A is web-similar to at most one idea in B and, vice versa, that every idea in B is web-similar to at most one idea in A . This means that web-similar ideas in A and B are connected by a bijection f from $\text{DOM}(f)$, some subset A' of A , to $\text{RAN}(f)$, some subset B' of B , in the sense that for all $a \in A'$ and $b \in B'$ it holds that a and b are web-similar if and only $b = f(a)$. In addition to this, on account of the fact that being web-similar means being related in the same way to other web-similar ideas, it cannot fail to hold that whenever $\langle a, a' \rangle \in R$, then for web-similar $f(a)$ and $f(a')$ also $\langle f(a), f(a') \rangle \in S$ and whenever $\langle a, a' \rangle \notin R$, then for web-similar $f(a)$ and $f(a')$ also $\langle f(a), f(a') \rangle \notin S$. That is: $\langle a, a' \rangle \in R$ if and only $\langle f(a), f(a') \rangle \in S$. In other words, the bijection $f: A' \rightarrow B'$ must be an *isomorphism* from $\langle A', R' \rangle$ to $\langle B', S' \rangle$, a function which respects the binary relations R' and S' , where R' and S' denote the restriction of R and S to A' and B' , respectively. (Cf. Chang and Keisler 1990, p. 21).

Consider, by way of illustration, Figure 3, which depicts the webs-of-ideas $\langle A, R \rangle = \langle \{a_1, a_2, a_3\}, \{\langle a_1, a_2 \rangle, \langle a_2, a_2 \rangle, \langle a_3, a_3 \rangle\} \rangle$ and $\langle B, S \rangle = \langle \{b_1, b_2, b_3\}, \{\langle b_1, b_3 \rangle, \langle b_2, b_2 \rangle, \langle b_3, b_3 \rangle\} \rangle$. Observe that the ideas in $\langle A, R \rangle$ and $\langle B, S \rangle$ are web-similar on account of the bijection $f: A \rightarrow B$ defined by $f(a_1) = b_1$, $f(a_2) = b_3$, $f(a_3) = b_2$, for note that f is an isomorphism from $\langle A, R \rangle$ to $\langle B, S \rangle$: for all $a, a' \in A$ we have that $\langle a, a' \rangle \in R$ if and only $\langle f(a), f(a') \rangle \in S$.

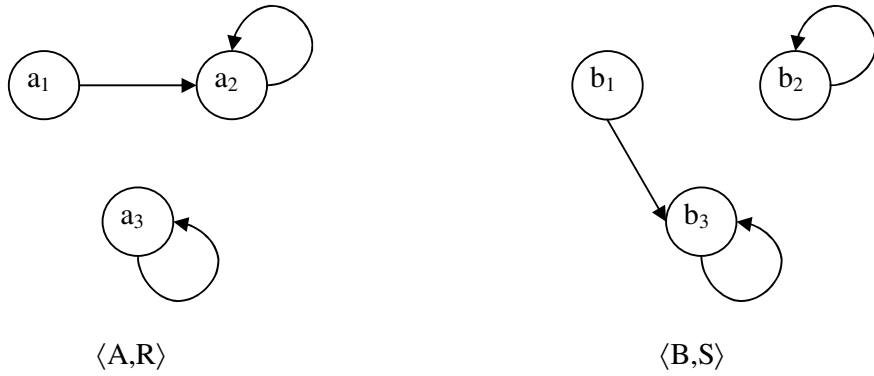


Figure 3

In Figure 3, the bijection f is unique in the sense that it is the only possible isomorphism between A and B . But this is not necessarily always the case. Thus in Figure 4 below, the webs-of-ideas $\langle A,R \rangle = \langle \{a_1, a_2, a_3\}, \{\langle a_1, a_2 \rangle, \langle a_2, a_1 \rangle, \langle a_2, a_3 \rangle, \langle a_3, a_2 \rangle\} \rangle$ and $\langle B,S \rangle = \langle \{b_1, b_2, b_3\}, \{\langle b_1, b_2 \rangle, \langle b_2, b_1 \rangle, \langle b_2, b_3 \rangle, \langle b_3, b_2 \rangle\} \rangle$ are connected by two isomorphisms $f:A \rightarrow B$ and $g:A \rightarrow B$, defined by $f(a_1) = b_1, f(a_2) = b_2, f(a_3) = b_3$, and $g(a_1) = b_3, f(a_2) = b_2$ and $g(a_3) = b_1$, respectively.



Figure 4

Of course, this situation can be disambiguated by embedding $\langle A,R \rangle$ and $\langle B,S \rangle$ in a larger context. Suppose that the webs of ideas $\langle A,R \rangle$ and $\langle B,S \rangle$ of Figure 4 are connected to respective sensory and motor systems S_A, S_B and M_A, M_B in that R and S are extended to relations R' and S' so that also $\langle S_A, a_1 \rangle \in R', \langle a_3, M_A \rangle \in R', \langle S_B, b_1 \rangle \in S'$ and $\langle b_3, M_B \rangle \in S'$, as illustrated in Figure 5. In that case, only (an extension f' of) the isomorphism $f:A \rightarrow B$ defined by $f(a_1) = b_1, f(a_2) = b_2, f(a_3) = b_3$ (and $f'(S_A) = S_B, f'(M_A) = M_B$) will 'survive'.

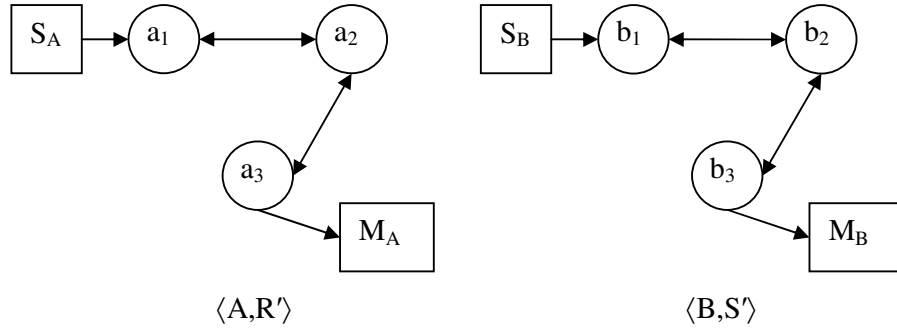


Figure 5

We have now analyzed web-similarity between ideas in webs-of-ideas $\langle A, R \rangle$ and $\langle B, S \rangle$ in terms of the existence of a bijection f from some subset A' of A to some subset B' of B which is also an isomorphism from $\langle A', R' \rangle$ to $\langle B', S' \rangle$, where R' and S' denote the restriction of R and S to A' and B' , respectively. Given this much, we may extend the concept of web-similarity and let it apply to full webs-of-ideas as well. Then we can distinguish pairs of webs-of-ideas that are web-similar in a – very – strict sense and pairs of webs-of-ideas that are only web-similar in a more rough sense. Strict web-similarity arises between webs-of-ideas $\langle A, R \rangle$ and $\langle B, S \rangle$ if the respective subsets A' and B' of A and B are non-proper subsets; that is: $\langle A, R \rangle$ and $\langle B, S \rangle$ are *strictly web-similar* if and only if there exists a bijection f from A to B which is also an isomorphism from $\langle A, R \rangle$ to $\langle B, S \rangle$. And non-strict, rough web-similarity arises between webs-of-ideas $\langle A, R \rangle$ and $\langle B, S \rangle$ if the respective subsets A' and B' of A and B are proper subsets. That is: $\langle A, R \rangle$ and $\langle B, S \rangle$ are *roughly web-similar* if and only if there exists a bijection f from some proper subset A' of A to some proper subset B' of B which is also an isomorphism from $\langle A', R' \rangle$ to $\langle B', S' \rangle$, where R' and S' denote the restriction of R and S to A' and B' , respectively. The above definition of strict web-similarity makes very high demands on candidate webs-of-ideas $\langle A, R \rangle$ and $\langle B, S \rangle$, since it requires that the webs-of-ideas involved actually have identical structures. For practical purposes such as communication between human subjects, however, less strict notions of web-similarity are presumably sufficient. (Note that we could actually make use of the subset-relation between different subsets of A in order to be able to distinguish different degrees of strictness.) Further research is required in order to decide how much less strict the notion of web-similarity may

become without communication turning into miscommunication. Nevertheless, the notion of strict web-similarity is important, since it describes a situation in which two different subjects S_1 and S_2 may fully realize the ideal of intersubjectivity, in that for every idea in the web-of-ideas of S_1 there is a web-similar idea in the web-of-ideas of S_2 , and vice versa.

Observe that the above definition gives the relation of strict web-similarity the desired properties of an equivalence relation. First, web-similarity is reflexive: a structure $\langle A',R' \rangle$ cannot fail to be connected to itself by the identity isomorphism f defined by $f(a) = a$ for all $a \in A$. Second, web-similarity is symmetric: if a structure $\langle A',R' \rangle$ is connected to a structure $\langle B',S' \rangle$ by the isomorphism f , then the structure $\langle B',S' \rangle$ is connected to the structure $\langle A',R' \rangle$ by the inverse isomorphism f^{-1} defined by $f^{-1}(f(a)) = a$ for all $a \in A$. And third, web-similarity is transitive: if a structure $\langle A',R' \rangle$ is connected to a structure $\langle B',S' \rangle$ by the isomorphism f and the structure $\langle B',S' \rangle$ is connected to a structure $\langle C',T' \rangle$ by the isomorphism g , then the structure $\langle A',R' \rangle$ is connected to the structure $\langle C',T' \rangle$ by the composition isomorphism $f \circ g$ defined by $f \circ g(a) = g(f(a))$ for all $a \in A$.

It should be noted that web-similarity is a holistic notion. The above formal definition shows that it is perfectly possible to define the notion as a relation between webs-of-ideas, but that does not mean that it corresponds to a straightforward relation between individual ideas within such webs. Recall that the formal definition was based on the intuition that ideas of possibly different subjects are web-similar if and only if they occupy ‘the same’ position within their own webs-of-ideas: within their own webs-of-ideas, web-similar ideas have to be related in the same way to all other web-similar ideas. Now, if the latter is taken seriously as a definition of web-similarity we run into the problem that the *definiendum* returns in the *definiens*. Since it lacks a sound basis, it cannot be understood as an inductive definition and must be judged circular. Consequently, checking whether the definition applies will get us caught in an infinite loop. For suppose that, given two webs-of-ideas $\langle A,R \rangle$ and $\langle B,S \rangle$ plus a bijection f from some $A' \subseteq A$ to some $B' \subseteq B$, we want to determine whether some $a_i \in A'$ is actually web-similar to $f(a_i) \in B'$. Then we have to check for each a

$\in A' - \{a_i\}$ whether a is web-similar to its counterpart $f(a) \in B' - \{f(a_i)\}$. But if we want to check for some particular $a_j \in A' - \{a_i\}$ whether it is actually web-similar to $f(a_j)$ in $B' - \{f(a_i)\}$, we will have to check for each $a \in A' - \{a_j\}$ whether a is web-similar to its counterpart $f(a) \in B' - \{f(a_j)\}$. However, since we can be sure that a_i is one of the elements of $A' - \{a_j\}$, this will bring us back to our initial problem: we will also have to check for $a_i \in A' - \{a_j\}$ whether it is web-similar to $f(a_i) \in B' - \{f(a_j)\}$...

The concept of web-similarity described above touches upon more possibly problematic issues. Recall that ideas of possibly different subjects are web-similar if and only if they occupy 'the same' position within their own webs-of-ideas: within their own webs-of-ideas, web-similar ideas have to be related in the same way to all other web-similar ideas. Thus web-similarity in effect reduces similarity of ideas to isomorphism of relations between ideas. Now, this raises the important question whether isomorphism of relations between ideas can be based on a publicly available criterion. For it seems that if the ideas of which webs-of-ideas consist are private entities, then webs-of-ideas must be private entities as well. And if this entails that the concept of web-similarity cannot be based upon public criteria, then the concept does not contribute to the solution of anything and we are again stuck in the private realm of the subject, unable to meet the intersubjectivity objection. So if web-similarity is the solution to this problem, then there should be some way of making the relations between ideas public. As was made clear above, there is nothing to make the *content of ideas* public and this was argued to lead to the intersubjectivity objection. However, we also suggested that the content of an idea can be identified by its position in the web-of-ideas, which is determined by the way in which the idea is related to other ideas (and eventually, via the sensory-motor system, to the outside world). And in addition to this we would like to argue that, in contrast to the mysterious privacy of a primitive concept of *content of ideas*, there *is* something with the help of which it is possible to make *relations between ideas* within webs-of-ideas public: viz., signs. Ideas owe their content to the way in which they are related to one another and to the outside world, so that the content of an idea is captured by its position within the web-of-ideas. And this position within the web-of-ideas can be made public with the help of the signs that stand for the ideas. Thus the intersubjectivity objection can be met

via the detour that content is determined by relations which can be made public with the help of signs.

Now, if signs are capable of making the relations between ideas public, then how does this work? Relations between ideas can be made public by some subject through its uttering concatenations of signs that stand for related ideas. These concatenated signs reflect the relations between the corresponding ideas and thereby publicize the private content that is captured by these relations. By way of an example, consider the idea TREE. Suppose that two subjects A and B have webs-of-ideas $\langle A, R \rangle$ and $\langle B, S \rangle$ with respective ideas $TREE_A$ and $TREE_B$, both expressed by the sign 'tree'. These subjects want to find out whether these ideas $TREE_A$ and $TREE_B$ are similar. The only way for A and B to test the similarity of $TREE_A$ and $TREE_B$, besides pointing to trees, is by investigating the relations of these ideas to other ideas such as, for example, the ones expressed by the signs 'leaf', 'stem' and 'branch'. These relations can be determined by checking each other's response to concatenations of the relevant signs, such as "tree has stem" and "branch has leaf". When a subject utters such a concatenation of the signs that stand for the ideas TREE and STEM, this makes public the fact that in this subject's web-of-ideas the ideas TREE and STEM are related and this in turn makes public that the ideas TREE and STEM have related content. So what can be tested using concatenated signs is whether the idea $TREE_A$ of subject A is related to other ideas in the web-of-ideas $\langle A, R \rangle$ in the same way as the idea $TREE_B$ of the subject B is related to other ideas in the web-of-ideas $\langle B, S \rangle$. In this way, isomorphism of (relevant parts of) different web-of-ideas can be revealed.

The previous description of concatenated signs might raise objections. Thus it may be noted that the above TREE-LEAF-STEM relations in webs-of-ideas are similar to what linguists call lexical networks.¹ Further research should determine the exact extent of this similarity, but it does give rise to the objection that human subjects rarely reflect on paradigmatic relations such as those between TREE, LEAF and STEM. Instead, they are more concerned with more contingent² relations such as those

¹ This was observed by prof. dr. H.E. de Swart.

² More contingent in the sense that not all cars are blue and not all buildings are high. Such relations are probably specific to some situation in which the subject may find itself or and have a less necessary character than the relation between TREE and STEM.

between CAR and BLUE or between BUILDING and HIGH: “car is blue”, “building is high”. A possible response to this objection is that the only difference between the paradigmatic relations which are made public by the concatenated signs “tree has stem” and “branch has leaf” on the one hand and the relations that are made public by the concatenated signs “car is blue” and “building is high” on the other, is one of permanence. The second type of relation might simply have a more transient appearance in the web-of-ideas than the first, but still require the same technique of being made public, viz., it is not the content of an idea directly but the relations that the idea has with other ideas that are made public, and because these relations – the position that the idea occupies within the web-of-ideas – capture its content, the private content of the idea is indirectly made public whenever the relations are made public with the help of concatenated signs.

In the previous paragraph it was claimed that the only way for two subjects A and B to test the similarity of their ideas $TREE_A$ and $TREE_B$, *besides pointing to trees*, is by investigating the relations of these ideas to other ideas. The italicized part refers to an important element: the anchoring of webs-of-ideas in reality. Although the activation process of ideas will be treated in more detail below, we will now go into its relevance to webs-of-ideas and the way in which it can be used as a benchmark for making the webs-of-ideas of different subjects converge. The activation process of ideas refers to the causal connection between reality and ideas: specific situations in reality activate specific ideas by triggering the subject’s sensory systems. From these sensory systems, signals proceed to certain ideas. For example, real trees tend to activate the idea TREE within human subjects. Note that such situations in reality are public, in that they are accessible to other subjects. This means that the activation process of ideas, which is embedded in webs-of-ideas, can be exploited in a similar way as the relations between ideas. Given a specific situation in reality, for example one in which a tree is prominently present, a subject may react by uttering ‘tree’ (the sign that stands for the idea TREE that is activated by the specific situation in reality). Now, another subject may see the same specific tree-containing situation in reality and hear the utterance as well, but fail to activate the idea for which the sign ‘tree’ stands in the situation at hand. Note that in this way it is possible for two subjects to compare and, moreover, adapt the activation process of certain ideas. So the activation process of ideas functions as an anchor that may keep webs-of-ideas both in

line with reality and similar to the webs-of-ideas of other subjects. There are, however, some difficulties with this account. Wittgenstein observed, as noted above, that it is impossible to learn a language without assuming certain innate rules or agreements. These difficulties will be discussed below. For now, it may be noted that the anchoring of webs-of-ideas within reality supplies another publicly available criterion, in addition to web-similarity, that can be invoked in an empiricist mentalist account of intersubjectivity for some specific ideas.

Observe that Wittgenstein claimed that in order to determine the information for which signs stand, one has to look at the way in which signs are used in language games. For example, in order to determine the information for which the sign 'tree' stands, we should look at the way in which this sign is used: thus the information for which this sign 'tree' stands is simply identified with the way in which it is used. The use of the sign 'tree' would probably include (1) *the other signs with which it occurs in sentences*; and (2) *the situations in reality in which human subjects tend to use the sign*. This description of the use of the sign 'tree' is reminiscent of the way in which the web-similarity of ideas was tested, for recall that in the above discussion of the sign 'tree', (1) the only property of 'tree' that seemed relevant from an outside, public perspective was how the sign 'tree' *occurred in combination with other signs* such as 'leave', 'stem' and 'branch'; and (2) it was suggested that two subjects, in an attempt to settle their dispute about the sign 'tree', put themselves in real situations in which trees are present in order to compare how their webs-of-ideas, and more specifically their TREE ideas, were anchored to reality by the activation process of ideas, so that here they were *using the sign 'tree' in relation to specific situations in reality*. In short, it seems that a more or less obvious way of making the Wittgensteinian assumption that 'meaning is use' more explicit is in keeping with our concepts web-of-ideas and web-similarity if these are considered from a public perspective, because the properties that can be derived from that assumption are also the ones exploited by web-similarity.

Web-similarity relies heavily upon the concept of webs-of-ideas and hence on the relations between ideas within them. However, the exact nature of these relations is presently quite unclear. Though further clarification will require future research, this does not necessarily render the present proposal useless. First, the concept of a web-of-ideas is a plausible one, because ideas within, for example, human subjects

intuitively have relations between them and, second, the proposal only depends in a very general way on relations between ideas that can be made public using concatenated signs. Until the exact nature of the webs-of-ideas is better known, we shall have to cope with the overly simplified view of a web-of-ideas as a directed graph (i.e., a set of nodes connected by arrows) and the distortions that may result from this simplification. On the other hand, if the required (neuro-)psychological research will have been carried out, the structure of actual webs-of-ideas revealed by it may actually turn out to be too complex to reason about, whereas the present simpler model suffices for achieving the goals of this master's thesis. Two additional possible properties of actual webs-of-ideas that are lacking from the ones discussed above are vagueness and ambiguity. Individual ideas in webs-of-ideas are described above as discrete and unambiguous nodes in the web, while in reality they are often vague and ambiguous. Another difference between actual webs-of-ideas and the ones discussed above concerns the relations between ideas. In our hypothesized webs-of-ideas the relations between two distinct ideas are represented by a single arrow between the corresponding nodes, suggesting a fundamental difference between ideas (the nodes) and their relations (the arrows). This, however, might turn out to be another simplification made for expository purposes, for our webs-of-ideas were introduced as a model with the sole function of showing that web-similarity can be successfully invoked in meeting the intersubjectivity objection against empiricist mentalist semantic theory.

2.2.2 The Second Problem

In the previous section it was outlined how the notion of web-similarity can be used in an empiricist mentalist account of the possibility of intersubjectivity, in spite of the fact that according to empiricist mentalist semantic theory subjects always have distinct subjective (and hence private) ideas. That is: the previous section showed that it is *theoretically* possible for the intersubjectivity condition to be met in an account based on web-similarity. However, the very fact of this theoretical possibility does not tell us anything yet about *how* this condition can be met in practice. Wittgenstein showed that the answer to this question is not obvious at all, because language cannot develop out of nothing; it requires some basis. The present section tries to answer the

question how different subjects may come to arrive at signs having the stand-for relation to web-similar ideas.

It is convenient to keep the following distinction in mind. From the perspective of the subject, there are two relations that involve ideas.¹ The first one is the stand-for relation, which holds between signs and ideas. The second one is the activation relation, which holds between situations in reality and ideas. Note that this does not mean that every idea is actually involved in both relations. Sometimes only the stand-for relation is present, sometimes only the activation process, and sometimes both relations are absent.

Another important distinction to be made is the one between the concept of 'subject' and the concept of '*human* subject'. The interdisciplinary approach of the present master's thesis entails that it employs the phrase 'subject' to refer to more than just *human* subjects. Its proposals may apply to a broader class that includes, for example, computer subjects.

The previous section already sketched a preliminary solution to the problem of how different subjects may arrive at having signs stand for web-similar ideas, but that solution only solved part of the problem. It came down to the claim that the combination of a common outside world, from which all subjects develop their ideas, and common instruments of idea extraction, with which all subjects develop these ideas, results in a situation where subjects have more or less the same ideas. Note, however, that this solution only talks about similar ideas but does not deal with the other requirement of the intersubjectivity condition, viz., that these 'more or less the same' – i.e., web-similar – ideas are connected to the same signs by the stand-for relation. Below three phases will be described which jointly constitute a process which may serve as an explanation of how the intersubjectivity condition can actually be met in the framework of an empiricist semantic theory.

Imagine a group of (fully developed) subjects which do not communicate or speak any language yet. Now, how could the intersubjectivity condition possibly come to be met by such a group? Recall that from the perspective of empiricist mentalist semantic theory, the condition requires that all subjects have similar webs-

¹ Stokhof (2003, pp. 16-23) discusses the relations between thinking, language and reality quite thoroughly.

of-ideas that are anchored to reality in a similar way. Observe, moreover, that learning a language in the given situation does not involve a learning process in the sense in which human subjects learn a second language at school. Rather, the situation is one in which the basis for language, intersubjectivity, must be constructed from almost nothing. This is reminiscent of the problem of initiating a language that was observed by Wittgenstein, who showed that without a common 'form of life', i.e., without some agreement or common ground between subjects, it is impossible for a subject to understand anything another subject tries to communicate, which means that the intersubjectivity condition cannot be met. Let us therefore assume that the imagined group of subjects has a common nature which possibly includes shared goals such as survival, collecting food and fear of predators, but also shared preferences, among other things. If we further assume that these subjects find themselves in a common outside world and dispose of common instruments of idea extraction, the preliminary solution sketched before will account for the possibility that the subjects may arrive in a situation where they have more or less the same ideas.

The following three phases make up a process that may provide an account of how our imagined subjects may meet the intersubjectivity condition.

Phase 1: Because the subjects have a common outside world and similar idea-extraction instruments they will display a shared tendency to extract more or less similar ideas in the sense that these similar ideas are activated by similar situations in reality and thus similarly anchored to reality as a result of their activation process. Thus far, however, they may lack any 'agreement' on the question which sign is to stand for which idea.

Phase 2: Because the subjects have more or less similar relations with reality and frame this reality in more or less similar ways (as a result of the previous phase), they may agree, through interaction based upon their shared goals and preferences, on using certain signs for these web-similar ideas. This interaction is an important factor. The shared goals and preferences make it possible to stimulate agreement and suppress dissension. This explains how the subjects attain roughly web-similar ideas, although they may still differ on a more detailed level.

Phase 3: The rough web-similarity of the subjects' ideas serves as the basis for further interaction between the subjects which, through convergence, may lead to more and more web-similarity between their ideas. And as the webs-of-ideas of the

subjects start to correspond more closely, their communication will be facilitated, thus increasing the speed with which new ideas are able to spread among different subjects. In the long run this leads to more fine-grained webs-of-ideas, reflecting the increased expressive power of the group's language.

A nice metaphor for this process is the way in which a company gets started. A starting company meets a similar problem: the company needs money in order to be able to offer its services, but it needs to offer its services in order to be able to earn the money. This deadlock is overcome by initial capital from a bank which is used for offering some limited services, which bring in some extra money, which may result in improved services, which yield extra money, etc., etc. Now, just as a starting company relies on a jump-start empowered by initial capital supplied by a bank, a language in the making relies on a jump-start provided by a common outside world, common instruments of idea extraction, as well as shared goals and preferences.

The above three phases make up a picture that suggests that after a jump-start, interaction between subjects is responsible for the convergence of the web-similarity of ideas and the signs that stand for these ideas. How does this *convergence through interaction* work? It should be observed that the convergence assumes that there is already some primitive form of communication, as well as some basic correspondence between subjects with respect to the web-similarity of their ideas and the signs that stand for these ideas. Now, in activities such as collecting food or protecting fellow subjects, the signs and the activation process of ideas of the subjects become publicly available and therefore can be tested on their correspondence with the signs and the activation process of ideas of other subjects. There are several factors that may be held responsible for the next step: the convergence. Note, by the way, that it is *not* important to which point the convergence proceeds – what matters is *that* it proceeds, for intersubjectivity only demands similarity and does not prefer one web-of-ideas to another. A first factor is authority. It is reasonable to assume that younger subjects try to imitate older subjects such as their parents. This will cause younger subjects to adapt their signs, ideas and stand-for relations to those of the authority. A second factor is hierarchy. The group of subjects may be hierarchically ordered, so that subjects lower in the hierarchy tend to adapt to subjects higher in the hierarchy. A third factor is average. A subject may adapt to what is most common among other subjects because this leads to the most effective communication. Together, these

factors are responsible for the convergence of the signs, ideas and stand-for relations in a process that clarifies how an imaginary group of subjects may actually implement the concept of web-similarity that was described in the previous section.

A possible objection to the above analysis of subjects' converging signs, ideas and stand-for relations is that in mathematics the concept of convergence also applies to two or more lines that will only meet one another in infinity. With respect to the above analysis, this would mean that the ideas, signs and stand-for relations of the subjects would be actually fail to be the same until an infinite amount of time would have passed. Of course, this means that they will never be the same in practice, and this entails that intersubjectivity will never be obtained.

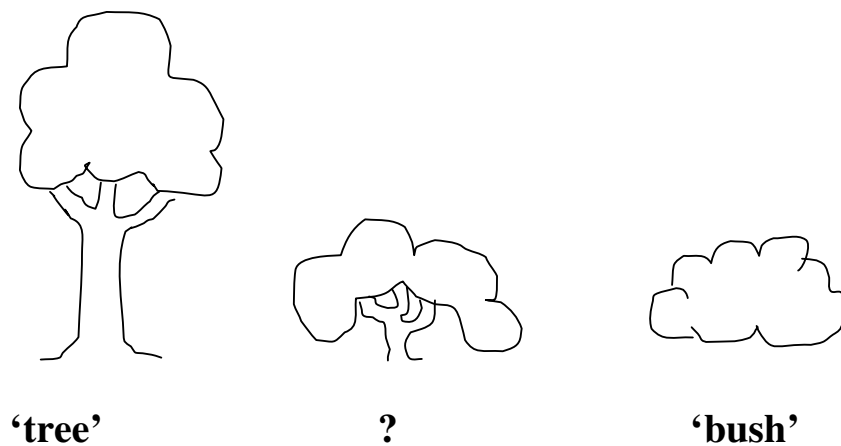


Figure 6

The problem with this objection is that it is based on the assumption that intersubjectivity can only be achieved when subjects have *exactly* the same webs-of-ideas, signs and stand-for relations. However, it was already suggested above that for actual communication rough web-similarity may already suffice, in that strict web-similarity is not (always) required. In order to show this, we may take two signs, say 'tree' and 'bush', as well as their respective corresponding ideas TREE and BUSH, and show random subjects pictures of trees and bushes. The subjects label the pictures with either 'tree' or 'bush' (and thus TREE or BUSH) whenever they believe the picture displays a tree or bush, respectively. Now, probably in most cases the subjects will agree with each other on the question whether the picture displays a tree or a bush, but in some cases the subjects are likely to disagree. The cases where the subjects do not agree on whether some picture shows a 'tree' or 'bush' illustrate that in actual practice subjects indeed fail to have exactly the same ideas; in many cases the ideas of subjects

are apparently slightly different. This not only means that, in practice, the intersubjectivity condition is hardly ever fully met, but it also supports the suggested process of convergence in that it indeed results in signs and ideas that are not exactly the same, just as its opponents suggest. It seems, then, that intersubjectivity and miscommunication come in amounts and degrees and that, moreover, the amount of intersubjectivity is more or less inversely proportional to the degree of miscommunication.

Chapter 3 Self-Applicability

This chapter takes up the second sub-question of the present master's thesis, viz., is it possible to meet the self-applicability objection against empiricist mentalist semantic theory within an interdisciplinary approach, and if so: how? The chapter consists of two sections. The first one gives a detailed description of the self-applicability objection against empiricist mentalist semantic theory. Several versions of this objection will be treated and confronted with Locke's thoughts about application criteria for ideas. Moreover, two problems will be argued to lie at the heart of the self-applicability objection, in the sense that incorporating a solution to these problems would make the resulting theory less vulnerable to the self-applicability objection. Some existing proposals for solving these problems will be discussed and evaluated. The second section of this chapter addresses the two central problems in the context of the existing proposals for solving them, proposes some further solutions, and investigates the consequences of these solutions for empiricist mentalist semantic theory in general.

3.1 The Self-Applicability Objection

In the first chapter we already shortly described the self-applicability objection against empiricist mentalist semantic theory, which will be discussed more thoroughly below. The present section consists of five parts. The first part explains the notion of application (or activation) criteria against the background of the activation process of ideas. The second part explains the self-applicability objection and introduces two derived problems that seem fundamental to it. The third part shows that the objection does apply to the work of the prominent empiricist John Locke. The fourth and the fifth part present and evaluate proposals for solving the derived problems.

3.1.1 Application Criteria

The self-applicability objection concerns the activation process of ideas. As was already made clear in the first chapter, this process takes place in a situation where ideas of a subject have already been developed and are, hence, present in the subject. In such a situation the process consists of the activation by situations in reality of such ideas that are already present. For example, consider some subject that has already

developed the idea TREE. Then the activation process involves how a tree in reality activates the idea TREE within this subject. In this situation one may distinguish the criteria that situations in reality must meet in order for the subject to apply the idea TREE to them. These criteria are usually called the *application criteria* of an idea. In this section we will also use the phrase *activation criteria* to refer to them, for although *application* is the most common term in the literature, *activation* is the term preferred by the present master's thesis in view of the fact that, as will soon become clear, in the context of empiricist mentalist semantic theory activation may well be a more appropriate concept than application.

Several authors consider such application criteria an essential component of every semantic theory. Bennett (1971, p. 24) discusses, in relation to Locke's theory, application criteria along the following lines:

If we classify things by checking them against 'patterns' or samples, how do we know which features of a given sample are relevant? Someone wonders whether to describe x as 'red', and he has established, somehow, that S is a sample of red and that S resembles x . But the resemblance is not total: in fact (though he is not in a position to say this), S is red, square, made of tweed, rough to the touch, and smelling faintly of lavender; while x is red, round, made of silk, etc. How can he select from among all the features of S the ones which are relevant to his problem of whether to describe x as 'red'?

Stokhof (2003, p. 39) also addresses the issue:

What in an idea determines whether it does or does not apply in a given situation? An idea, formed out of certain sensorial impressions, should also be applicable to new cases, and this application can be correct, but also incorrect. What are the criteria for correct application of an idea, and how do these criteria bear to the idea itself?¹

3.1.2 The Objection and Two Derived Problems

Opponents of empiricist mentalist semantic theory argue that the necessity of having application criteria of ideas, which are supposed to determine whether some 'observable similarity' with some situation in reality actually obtains, force

¹ My translation of: 'Wat in een idee bepaalt of het in een gegeven situatie wel of niet van toepassing is? Een idee, gevormd uit bepaalde zintuiglijke impressies, moet ook op nieuwe gevallen kunnen worden toegepast, en die toepassing kan correct, maar ook incorrect zijn. Wat zijn de criteria voor correcte toepassing van een idee, en hoe verhouden die criteria zich tot de idee zelf?' (Stokhof 2003, p. 39).

empiricists to assume that ideas have a self-applying character. The concept of a self-applying idea entails that the application criteria of such an idea are intrinsic to the idea itself. And because such self-applying ideas are considered highly implausible entities, the opponents conclude that empiricist mentalist semantic theory must be incorrect.

It may be observed that the self-applicability objection actually has two parts: the first part argues that empiricists must assume ideas to be self-applying; and the second part argues that self-applying ideas invalidate empiricist mentalist semantic theory. In the present subsection we will go further into various details of the objection.

In order to clarify the first part of the objection, that argues that ideas must be self-applying, we consider an argument by Wittgenstein as explained by Stokhof (2003, p. 48-49). This argument shows that if ideas are not self-applying, then the process of applying an idea to some object in reality results in an infinite regress. According to Stokhof, Wittgenstein assumes that ideas are mental pictures. Moreover, pictures are objects too, and this means that before some object can be said to be a picture, it needs to be interpreted as such. The next step involves testing whether the picture's application criteria apply to the object in reality by determining whether the picture is observably similar to the object at hand, but before this can occur it must be determined how the picture itself will be interpreted, for most pictures are interpretable in many different ways. Wittgenstein uses the example of a picture of a cube with a specific method of projection (such as the famous Necker cube below) to show that one picture can correctly be applied to different situations in reality.

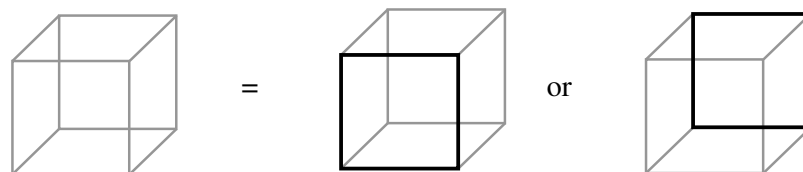


Figure 7

According to Wittgenstein, the interpretation of the picture is itself a picture: a method of projection. And this leads to an infinite regress because the picture of the method of projection must be interpreted too, which involves another interpretation,

etc. This argument does not only apply to spatial objects such as a cube but can be applied in general. Every object (whether it is a linguistic object such as a word or a non-linguistic object such as the idea of a tree) requires interpretation before it can be used as a picture. According to empiricist mentalist semantic theory, this interpretation (an idea) is itself a (mental) object and thus requires interpretation too, before it can be used. But this means that another mental object is required that needs interpretation, etc. Michael Ayers also shortly considers this argument (Ayers, 1991 pp. 248-249), although he believes it does not apply to Locke's theory.

The previous above shows that interpreting in order to determine whether an idea is observably similar to some object in reality leads to an infinite regress of interpreting interpretations. The only way to avoid this is by assuming that ideas are self-applying. This means that an idea does not require interpretation at all, but mysteriously pops up whenever an object in reality satisfies its application criteria. But, according to various opponents of empiricist mentalist semantic theory, self-applying ideas are not very plausible and this brings us to the second part of the objection that uses the premise of the implausibility of self-applying ideas to conclude that empiricist mentalist semantic theory must be incorrect.

According to Wittgenstein, self-applying ideas are implausible because ideas are picture-like and pictures, as we use them in ordinary life, cannot be self-applying. They often have several different possible applications – and how can a picture 'know' which of its various applications is the correct one?

Another argument why self-applying ideas are implausible is provided by Stokhof:

So it seems that in a mentalist explanation of the meaning of language we have to appeal to the intrinsically 'meaningful' character of ideas. But this results in a theory that has little explanatory power: it explains the meaning of language in terms of the meaning of ideas.¹

Stokhof also mentions an argument from Wittgenstein that tries to point out that if ideas have a self-applying character they do not have to be observably similar to the

¹ My translation of: 'Het lijkt er dus op alsof we in een mentalistische uitleg van betekenis van taal een beroep moeten doen op het intrinsiek 'betekenisvolle' karakter van ideeën. Maar dat resulteert in een theorie die weinig verklarende kracht heeft: het legt betekenis van taal uit in termen van betekenis van ideeën.' (Stokhof 2003, p. 40).

objects in reality to which they apply. This is because observable similarity makes it possible to put the application criteria of an idea outside the idea itself, but when the application criteria are intrinsic to the idea the observable similarity, which is necessary for having application criteria outside the idea itself, is not necessary anymore. But if this is the case, that is:

[i]f we keep in mind the possibility of a picture, which, though correct, has no similarity with its object, the interpolation of a shadow between the sentence and reality loses all point. (Wittgenstein 1958, p. 37).

This means that the ideas are 'idle wheels' in the theory that are useless as mediators between sentences and reality, for their self-applying role can easily be taken over by the sentences themselves.

Jonathan Lowe also recognizes the first part of the above self-applicability objection, which argues that ideas must be self-applicable. He notes the following about one of Locke's application criteria of ideas, that of observable similarity:

[T]o know what sort of thing we are confronted with, we compare it with a pattern in our mental pattern-book. But the patterns [...] are *themselves particulars*, so how do we recognise *them* as being of a certain sort or kind? How do I recognize the abstract general idea with which I am comparing a newly encountered object as being an abstract general idea *of a man* (as opposed to one *of a horse*, say)? It looks as though Locke's theory of recognition generates a vicious infinite regress. (Lowe 1995, p. 164)

Lowe does not recognize the second part of the objection, however, which argues that empiricist mentalist semantic theory is incorrect because it is forced to assume self-applying ideas. In fact, Lowe believes that self-applying ideas constitute the very solution to the infinite regress. He states:

Locke, it seems, ought to say that the 'matching' of newly encountered objects to abstract general ideas is similarly an 'automatic' process, unmediated by further processes of 'recognition'. (Lowe 1995, p. 165)

In other words: the application criteria of ideas do not lie outside the ideas themselves, which would involve a process of recognizing the ideas, but are intrinsic to the ideas themselves and thus happen automatically. We will return to Lowe shortly, when we will discuss proposals for meeting the self-applicability objection against empiricist mentalist semantic theories.

All in all, we can summarize the various versions of the self-applicability objection as follows. Ideas must be self-applying, because non-self-applying ideas

require external criteria for determining when they are applicable, which leads to an infinite regress. But self-applying ideas are implausible and lack explanatory power, so that empiricist mentalist semantic theory, unable as it is to explain how ideas apply to reality, must be incorrect.

There are two fundamental problems that underlie the above objections, in the sense that solving both of them would make empiricist mentalist semantic theory immune to the self-applicability objection.

The first underlying problem concerns the nature of ideas. Wittgenstein assumed that they are pictures and this assumption was shown to lead to problems. The question is whether ideas can be assigned an alternative nature that is able to prevent these problems from arising. Stokhof also recognizes that clarifying the nature of ideas could avoid the self-applicability objection:

There is thus room, in principle, for an alternative view [...] which makes the opposite assumption: there are ‘self-applying’ mental contents.¹

The second underlying problem concerns the way in which application criteria function, or, in different terms, how ideas get activated. In the introductory first chapter this was referred to as the activation process of ideas, and it is this process that requires elucidation.

3.1.3 The Objection Applied to Locke

The previous subsection discussed some variants of the self-applicability objection against empiricist mentalist semantic theories. In the introductory chapter it was announced that Locke’s *An Essay Concerning Human Understanding* will be used as the main representative of empiricist mentalist semantic theory, and Locke has in fact already been mentioned a couple of times above. This subsection will further explore the relation between the variants of the objection and Locke’s empiricist mentalist semantic theory and show that the objections are indeed valid objections against it.

The first part of the objection focused on the criteria for applying some idea to a situation in reality. Wittgenstein and Stokhof claimed that application criteria for ideas are external to those ideas and, moreover, that they are determined by

¹ My translation of: ‘Er is dus ruimte, althans in principe, voor een alternatieve visie [...] die de tegenovergestelde assumptie maakt: er zijn ‘self-applying’ mentale inhouden.’ (Stokhof 2003, p. 49).

‘observable similarity’ of the idea with the situation in reality. The following passages from Locke’s *Essay* illustrate his view of application criteria and makes clear that it was indeed based upon similarity of the idea with some object in reality:

[T]o make abstract general *Ideas*, and set them up in the mind, with Names annexed to them, as Patterns, or Forms [...] to which, as particular Things existing are found to agree, so they come to be of that Species. (Locke 1690, Section III.iii.13)

Locke also says: “[b]ut it [is] evident, that Things are ranked under Names into sorts or *Species*, only as they agree to certain abstract *Ideas*” (Locke 1690, Section III.iii.6), and he adds that “what has a conformity to the abstract *Idea* the name *Man* stands for” (Locke 1690, Section III.iii.12). These passages suggest that application criteria of ideas rely upon ‘observable similarity’, although they do not explicitly mention the concept. Locke *does* explicitly mention the concept in relation to a particular class of ideas, when he says “[t]hat the *Ideas of primary Qualities* of Bodies, are *Resemblances* of them” (Locke 1690, Section II.viii.15). The phrases used most often by Locke are ‘agreement’ and ‘conformity’, and he does so without explaining in what sense he uses them. Combining these passages with those where he talks about ideas as images or pictures does make it plausible that Locke saw the application criteria of ideas are based on ‘observable similarity’.

The second part of the objection focuses on why self-applicable ideas are implausible. One of Wittgenstein’s arguments used the fact that ideas are picture-like to show this. Whether Locke indeed thought of ideas as image- or picture-like is therefore crucial to the validity of his argument. According to Michael Ayers this is not a settled question, he says, however, that with respect to some specific chapter (Ayers 1991, p. 46) “[i]t would be very difficult to read this chapter without taking recalled sensory ideas to be at least mental images, if not corporeal ones.” The fact that it is not a settled question due to Locke’s obscure standpoint, might already justify Wittgenstein’s assumption that ideas are picture-like.

3.1.4 Some Proposals

The two fundamental problems underlying the self-applicability objection against empiricist mentalist semantic theory concern the nature of ideas and the way in which ideas get activated. There have already been solutions proposed to these problems, some of which will be treated in the present subsection.

We will first go into a proposal for solving the problem of the activation process of ideas. It was already noted that the criteria governing this activation process are commonly called the application criteria of ideas in the literature. There it was also mentioned that Wittgenstein argued that Locke's idea of basing such application criteria on observable similarity leads to an infinite regress. How can this infinite regress be avoided? Observe that Stokhof and Wittgenstein already suggested the direction in which a solution should be sought. They said that the only way in which the infinite regress can be avoided is by letting ideas be self-applying, but they also argued that this is an implausible demand. However, a passage from Stokhof shows that he thinks that there is still some room for a solution along those lines. Note that the requirement that ideas be self-applying has consequences for the nature of ideas. Ideas that are picture-like, for example, were already shown to be incompatible with this requirement. This shows that the two problems are closely interrelated.

Jonathan Lowe also recognized that adoption of the concept of self-applying ideas could lead to a solution to the problem of the activation process of ideas in that his proposal concerning the nature of ideas does seem to actually adopt it. In the following passage he explains his proposal:

On Locke's behalf I think we could reply that the pattern-book model, though suggested by Locke's own remarks, is not in fact essential to his thesis that recognition is mediated by abstract general ideas. [...] If an analogy is sought, a better one than the pattern-book might be that of the automatic vending-machine, which 'recognizes' a coin because the coin matches a slot inside the machine: the machine does not require a further mechanism to 'recognize' *the slots*, so that it can 'compare' a particular slot with a particular coin in order to determine whether they match. Rather, the slot which matches is the one which the coin goes through. Locke, it seems, ought to say that the 'matching' of newly encountered objects to abstract general ideas is similarly an 'automatic' process, unmediated by further processes of 'recognition'. (Lowe 1995, pp. 164-165)

So, according to Lowe, ideas could have a nature that resembles the coin-slots in an automatic vending machine. The activation process of such ideas would also be similar to the activation process of coin-slots in such a machine. Lowe, however, leaves it at the above suggestion and does not investigate it any further.

A related proposal comes from Jerry Fodor. By way of defending his theory of 'The Language of Thought' against the self-applicability objection he says the following:

My point is that, though the machine must have a compiler if it is to use the input/output language, it doesn't also need a compiler for the machine language. What avoids the infinite regression of compilers is the fact that the machine is *built* to use the machine language. Roughly, the machine language differs from the input/output language in that its formulae correspond directly to computationally relevant physical states and operations of the machine: The physics of the machine thus guarantees that the sequences of states and operations it runs through in the course of its computations respect the semantic constraints on formulae in its internal language. (Fodor 1975, p. 66)

Fodor uses somewhat unfamiliar terminology, but his point comes down to the following. The tokens of the machine language do not require further interpretation (i.e., compiling). They are self-applying in the sense that the physical structure of the machine ensures a correct extension of tokens of the machine language. On account of the unfamiliar terminology the correspondence between empiricist mentalist semantic theory and Fodor's language of thought is less than obvious. It is thus far from clear to which of the two problems Fodor's proposal constitutes a solution, but what does become clear from Fodor's passage is that self-applicability is not as implausible as Wittgenstein and Stokhof want us to believe.

3.1.5 Do the Proposals Work?

The previous subsection outlined some existing proposals for solving the two fundamental problems underlying the self-applicability objection. The present subsection will shortly evaluate these proposals and sketch their possible shortcomings. We will first evaluate Lowe's proposal and then proceed to Fodor's suggestion.

One could object to the solution proposed by Lowe that a new difficult problem is introduced by removing the conscious aspect of recognizing and manipulating ideas. The development process of ideas as described by Locke is rendered obsolete because it explicitly requires conscious manipulation, as illustrated by the following passages:

[T]he Mind having got any Idea, which it thinks it may have use of, either in Contemplation or Discourse, the first Thing it does, is to abstract it, and then get a Name to it. (Locke 1690, Section II.xxxii.7)

[T]he Mind makes the particular Ideas, received from particular Objects, to become general; which is done by *considering* [my italics, BN] them as they are in the Mind. (Locke 1690, Section I.xi.9)

This has the consequence that the development process of ideas cannot be appealed to anymore for answering questions, so that a completely new account must be given. This is, of course, not a direct argument against Lowe's proposal, but it does show that Locke's formulations are at least inaccurate or that his account is simply incorrect and requires an update. The latter is what this master's thesis sets out to do. Hopefully, the account of the development process to be presented below is a satisfying alternative to Locke's account.

Both in Fodor's and in Lowe's proposal a physical structure is assumed to create the self-applying required to avoid the infinite regress that was signalled by Wittgenstein. In Lowe's proposal the physical structure is compared to an automatic vending machine. An automatic vending machine gets a coin inserted that is recognized because the physical structure of the automatic vending machine leads the coin into the correct coin-slot. While this does not seem strange for an automatic vending machine, it is not immediately clear how this can happen within a subject. Fortunately, a subject of which the idea TREE gets activated does not have to swallow a real tree which is then led to activate the idea TREE by the physical internal structure of the subject. But then, how *does* it happen within a subject? Obviously, further details are required for making the proposal more plausible.

The first half of the present chapter has given an overview of the most important components of the self-applicability objection against empiricist mentalist semantic theories. The notion 'application criteria' was explained, the self-applicability objection was explained with the help of Wittgenstein's and Stokhof's arguments and some proposals were discussed. These proposals notwithstanding, the self-applicability objection remains to bother empiricist mentalist semantic theory. To put an end to this, the proposals need to be worked out in more detail. We will now proceed to complement the proposals with new ideas in order to hopefully arrive at satisfactory solutions.

3.2 Meeting the Self-Applicability Objection

This section focuses on the way in which empiricist mentalist semantic theory may meet the self-applicability objection. A detailed description of this objection was given above, where the objection was partitioned into two underlying problems that need further clarification and some proposals from the literature were discussed. The

present section goes further into these problems and their solutions. The problem concerning the nature of ideas will be handled in Section 3.2.1 and Section 3.2.2 will discuss the problem regarding the activation of ideas.

3.2.1 The First Problem

The problem concerning the nature of ideas was introduced because a solution to this problem may have fundamental consequences for the self-applicability objection. We saw above that Wittgenstein argued that, given the assumption that ideas are picture-like or object-like, empiricist mentalist semantic theory is necessarily led into an infinite regress of interpreting interpretations. However, Stokhof recognized that there may be a way out of this regress, if one allows for the possibility that the nature of ideas may not be picture-like at all. According to Stokhof, empiricist mentalist semantic theory will be able to avoid the regress as soon as it allows its ideas to be of a self-applying nature, but Stokhof does not consider this a very plausible option. But leaving aside the issue of plausibility, it can be observed that Stokhof is right: if ideas are of a nature that allows them to be self-applying, the infinite regress is avoided – and this is exactly what Lowe’s proposal entails. Lowe suggests that ideas are like slots in an automatic vending machine, which “‘recognizes’ a coin because the coin matches a slot inside the machine” (Lowe 1995, p. 165). Wittgenstein’s argument does not seem to apply to slot-like ideas. The present section investigates whether ideas can be slot-like and whether such ideas indeed render Wittgenstein’s argument inapplicable.

Because Lowe’s automatic vending machine metaphor is so striking we will elaborate it in the present section. A detailed description may therefore be useful. An automatic vending machine delivers, when the correct amount of coins is inserted, some snack or tin can. According to Lowe, the mechanism that decides whether some idea is applicable to some situation in reality that serves as input to the subject’s senses has analogies with the mechanism in an automatic vending machine that decides whether some coin-slot is ‘applicable’ to some situation in which coins are inserted by some gourmet. The exterior of an automatic vending machine has an opening through which coins can be inserted. Making the analogy more detailed, this opening may be compared with the senses of a subject (although senses are usually far more complex). Whenever a coin is inserted into the opening of the vending machine

it finds the mechanism behind this opening developed in such a way that different types of coins travel different routes through the internals of the automatic vending machine until they fall into a coin-slot for that particular coin. Whenever this happens, the type of coin is said to be recognized. In an analogous fashion, signals from situations in reality enter the senses of a subject and the structure or mechanism behind these senses has developed in such a way that signals from different types of situations in reality travel different routes through the internals of the subject until they activate a slot-like idea. Whenever this happens, the type of situation in reality is said to be recognized. Note that the process of activating slot-like ideas will be of concern for attempts at solving the second problem, which will be treated elaborately in the next section. The present section focuses on the question whether ideas are of a slot-like nature.

The analogy with automatic vending machines should contribute to a clarification of the nature of ideas, but does this alternative approach indeed render Wittgenstein's reduction to an infinite regress inapplicable? A key feature of Wittgenstein's argument, as will be recalled from Section 3.1.2, is that it is based on the assumption that picture-like ideas are objects and thus require interpretation, just as objects in reality do. This key feature was shown to lead to the infinite regress of interpreting the interpretation, interpreting the interpretation of the interpretation, etcetera. However, note that this problem does not arise with slot-like ideas, simply because slot-like ideas, due to their self-applying nature, do not require interpretation at all. That is: ideas do not have to be recognized or interpreted before they can be applied. Instead, they are mysteriously activated in situations in which they are applicable just as the coin-slots are mysteriously activated in situations in which they are applicable. The point is that with automatic vending machines nobody finds this implausible, while with ideas a lot of people do.

To increase the plausibility of the above claims, we may consider the possibility of slot-like ideas in human subjects. While Lowe is probably right when he notes that “[h]ow such a model [analogous to an automatic vending machine] might be actually implemented in the human mind or brain is, however, more a matter for empirical psychology than for philosophical speculation” (Lowe 1995, p. 165), our considerations will be mere speculations that are not corroborated by any empirical evidence. Well, then: a slot-like idea within a human subject might be a network of

neurons that can get activated. The human brain is structured in such a way that whenever the human subject sees (for example) a tree, some such network of neurons gets activated. Note that it is not of any importance which particular network of neurons gets activated. This may differ from human to human, although some correspondence is to be expected in view of the fact that all human brains are probably wired up in roughly similar ways, both internally and in relation to the senses.¹ Using so-called neural networks, a highly simplified model of the human brain can be implemented on computers. Such neural networks can, for example, be used to recognize slightly different versions of (pictures of) the digits zero to nine. Neural networks are very similar to slot-like ideas and automatic vending machines in the sense that in the course of the initial training or learning phase, the structure of such a network is carefully adapted, so that in the subsequent recognition phase a specific node in the network will be activated whenever a (picture of a) certain digit is fed to its input. This process also illustrates that these specific nodes, which each correspond to a particular digit, are self-applying, for they are mysteriously activated whenever a (picture of a) certain digit is fed to the input of the network, without any further process of recognition taking place. This may serve to illustrate that slot-like ideas are not that implausible and have a role to play in scientific disciplines such as neuroscience and computer science.

Talking about other scientific disciplines: there is an additional argument which may convince believers in determinism that self-applying ideas are not as implausible as they might seem. Proponents of determinism will say that, just like all other phenomena in reality, the semantic phenomenon should eventually be reducible to causal rules. Now, an account in terms of self-applying slot-like ideas as provided above fits this doctrine much better than the account in terms of the object-like or picture-like ideas that was considered by Wittgenstein and Stokhof. For at some point in the reduction of the semantic phenomenon the supervision must be handed over from the human subject to the causal laws of determinism and self-applying slot-like ideas suggest how this transition may take place. Picture-like ideas, on the other hand, suggest exactly the opposite. They suggest a desperate perseverance of human

¹ For example, the eyes of humans are commonly connected to the backside of the brain.

subjects in keeping supervision over the interpretation process, resulting in an endless chain of interpretations, interpreting interpretations, interpreting interpretations of interpretations, etc.

As we saw above, Wittgenstein and Stokhof put forward two arguments against the concept of self-applying ideas, which led them to conclude its implausibility. We will return to these arguments here and formulate some counterarguments.

The first argument is due to Stokhof. He argued that self-applying ideas have a ‘meaningful’ character and concludes that this creates an unsatisfying situation, namely one in which the meaningfulness of signs is explained in terms of the meaningfulness of ideas. However, this argument seems to be founded on the claim that self-applicable ideas have a meaningful character, while it is actually unclear why self-applicable ideas would have to be meaningful. Note, for instance, that this would mean that the coin-slots in an automatic vending machine have a meaningful character too. (By the way, even if these coin-slots had a meaningful character, is not meaningfulness exactly what is being explained?) In short, nothing forces self-applying ideas to be meaningful, and the argument fails.

The second argument is due to Wittgenstein. He argued that self-applying ideas are idle wheels in semantic theory. Because self-applying ideas do not require further interpretation they do not have to be similar in any sense to the object in reality that they represent. But this means that they are useless as mediators between signs and reality, for their self-applying role can easily be taken over by the signs themselves. Since their function is unclear, self-applying ideas are idle wheels in the theory that can be left out entirely. But while self-applying ideas do not require further interpretation and thus can do without similarity to the objects in reality, their mediation between signs and reality is crucial for another reason. As Stokhof notes, ‘the relation between language and meaning is assumed to be conventional’.¹ This means that the relation between signs and reality has to be ‘stored’ somewhere, since it does not follow from the laws of nature as opposed to, for example, the non-conventional relation between smoke and fire. The storehouse of the relations between signs and reality is the subject and within this subject these relations are

¹ My translation of: ‘De relatie tussen taal en betekenis wordt verondersteld conventioneel te zijn’ (Stokhof 2001, p. 40).

mediated by self-applying ideas. So the ideas, far from being idle wheels in empiricist mentalist semantic theory, are required to maintain the relations between signs and reality in the first place. Issues concerning the activation process of ideas will be addressed more extensively in the next section.

3.2.2 The Second Problem

The second problem underlying the self-applicability objection is that of how ideas get activated or, put more traditionally, what their application criteria are. Although this subject was already touched upon above, there are still sufficiently many issues that should be treated more elaborately. The previous subsection argued that the nature of ideas is slot-like, rather than object-like or picture-like, and this made it possible to avoid Wittgenstein's infinite regress of interpreting interpretations. But now that the nature of ideas has changed so that their application criteria do no longer rely upon interpreting, there is need for an alternative approach to the application criteria of ideas that is consistent with their slot-like nature. The present section will address several aspects of the activation of slot-like ideas.

Before these new application criteria will be treated, some terminological issues need to be clarified. When ideas are related to reality, one finds that often terms such as 'refers to', 'denotes', 'represents', 'means' or 'applies to' are used. This may have supplied fuel to the self-applicability objection, for it may be noted that these terms have a conscious or interpretative character, which is exactly what the self-applicability objection relies upon. Locke himself contributed to this as well, because he often describes processes as if they proceed consciously. Thus Locke says the following about the abstraction process of ideas:

Children [...] frame an *Idea*, they give [...] the name Man. [...] Wherein they make nothing new, but only leave out of the complex *Idea* they had of *Peter* and *James*, *Mary* and *Jane*, that which is peculiar to each, and retain only what is common to them all. (Locke 1690, Section III.iii.7)

Given their slot-like nature, the term 'activation' might be more appropriate to describe the relation between reality and ideas, and the term may even prevent people from thinking of this process as an interpretational one. Correspondingly, we will from now on replace the term 'application criteria' by 'activation criteria' and employ the phrase 'is-activated-by' to characterize the relation between ideas and reality. The

relevant processes may very well be unconscious ones, as is also noted by Fodor (1975, pp. 70-71).

In the previous section the automatic vending machine was used as an analogy to the process of the activation of slot-like ideas. According to this analogy the internal mechanism or structure of the subject ensures that situations in reality activate the correct slot-like ideas. What this means is that the activation criteria of an idea are implicitly determined by the internal structure of the subject. It is reasonable to ask whether this is actually possible. How can the internal structure of some subject determine the activation of ideas? This possibility may also become clear by an appeal to the analogy with the automatic vending machine. Inside the machine there are all kinds of cogs, levers and springs which, together with the forces of nature and possibly the different weight and size of coins, make sure that the inserted coin gets into the correct coin-slot. For more clarification, the structure of the subject can also be compared with a riverbed that conducts, together with the forces of nature, the water along a specific route towards the river mouth, where the water flows into an ocean, sea or lake. The structure of the landscape determines whether the forces of nature conduct the water to the ocean, sea or lake. In these cases it seems obvious how the structure of the automatic vending machine or the landscape may guide the 'activation' of different 'ideas'. The exact structure responsible for the activation of ideas within a subject is not prescribed and depends on the nature of the subject. If the subject resides inside a personal computer, the structure responsible for the activation of specific ideas may consist of the electronics of the processor and motherboard together with the magnetic fields of the hard disk and the memory. When a human subject is concerned, the structure responsible for the activation of specific ideas is probably brain tissue, in all its complexity. It is important to notice that the exact nature of the subject is not of importance for the account, as long as it contains the structure required to activate the ideas. This does not mean, however, that the type of structure does not influence the properties of the relations, because, for example, the relations between types of coins and coin-slots within an automatic vending machine are probably much less flexible than the relations between types of situations in reality and ideas within a human subject.

The above analogies suggest that in general there are two ingredients required to create the relations under consideration: the forces of nature and the structure of

reality. It is obvious why the forces of nature are required, for without them there would be no change at all, but it may be less clear why the structure of reality is required. This has to do with the fact that the relation between ideas and reality is of a different kind than, for example, the relation between water and ice or between smoke and fire. The fact that smoke is a by-product of fire and that cold water eventually turns into ice involves relations that do not need to be stored somewhere, because they solely depend on the forces of nature. The relation between ideas and reality on the other hand is a conventional one, which does not solely depend on the forces of nature and thus requires embedding in a storehouse or carrier. This storehouse is exactly the structure of reality that was mentioned above. Whereas relations that depend on the forces of nature can be called *causal relations*, relations that depend on both the forces of nature and the structure of reality can be called *embedded causal relations*. Relations between ideas and reality are thus examples of embedded causal relations and therefore require a carrier: the subject. The previous seems to be in keeping with Lowe's distinction between natural and artificial relations (Lowe 1995, pp. 148-149), where Lowe's natural relations are analogous to causal relations and his artificial relations are analogous to embedded causal relations.

Above we have considered the process of how an idea is activated by reality. Because this is an embedded causal relation it must be stored somewhere in the structure of reality. A subsequent question, then, is how (part of) the structure of reality comes to be such that it stores these relations, for they obviously have not always been there. In the case of automatic vending machines the mechanisms which ensure that every coin falls in the correct coin-slot are built by humans, but how does this come about within subjects? How does the internal structure of a subject become such that it correctly activates ideas when the senses of the subject are presented with situations in reality? These questions will be answered more elaborately in a section on the development process of ideas in the next chapter, but we will also shortly discuss them here.

It was already noted that there is no prescribed specific implementation in the subject and this also holds for the particular way in which the internal structure develops. This particular development might even depend on the specific implementation in the subject. In the case of a subject that is embodied by a personal computer, the internal structure may be developed or programmed by humans, just as

the internal structure of automatic vending machines, but in the case of human subjects this cannot be the case. The development of the structures in humans is probably the joint product of evolution, education and experience. In that sense the analogy with automatic vending machines does not apply here, whereas the analogy with a river that flows to the sea does (to a certain extent, that is). Analogous to rivers that develop due to water flowing from higher to lower elevations and scraping the riverbed in the process, the experience of a human subject might 'scrape' the relations between ideas and reality into its brain tissue.

The processes involved above may seem rather automated and driven by forces of nature rather than the subject's own will. This may please determinists, but it may strike others (and maybe also some determinists) as being incomplete. For how does the process ensure that the *correct* relations get 'scraped' into, for example, the brain tissue of a human subject? Or, in other words, how does the process ensure that an idea ends up with the correct activation criteria? There is no guarantee that the above process results in a situation where trees in reality will actually activate the idea TREE within the human subject. For all we know, the process could end up making trees in reality activate the idea DOG, which would obviously be incorrect! It must be noted, however, that this objection is the result of some misunderstandings. It is true that the process does not guarantee the correctness of the developed relations but, as noted above, the development process of ideas that is responsible for the creation of these relations may be influenced by several factors, including education. Because correctness and incorrectness in such cases are determined by standards shared by all human subjects in a society, it is to be expected that education, which is governed by these human subjects, will influence the development process in such a way that the human subjects being educated will develop the correct relations.

There is an important issue related to the activation of ideas that needs to be considered. As was noted above, it is the internal structure of a subject that determines which slot-like idea will be activated by a certain type of situation in reality. This brings us to the issue of the content of slot-like ideas, which was already mentioned a couple of times in previous sections. What can be said to be the content of, for example, the slot-like idea TREE? What happens when the slot-like idea TREE is activated? Whenever the slot-like idea TREE is activated, the perceived situation in reality is such that it satisfies the activation criteria of this idea. And moreover:

whenever a slot-like idea is activated by a perceived situation in reality that satisfies the activation criteria of the idea, this is a consequence of the fact that the internal structure of the subject is such that it conducts the sensory input signals to that particular slot-like idea via its activation process for that particular slot-like idea. And this is valuable information for the subject that is the owner of the slot-like idea, for on the basis of the activation of that idea the subject knows in what situation in reality it finds itself, so that it may make certain choices based on this knowledge. Hence the conclusion must be that it is the activation of a slot-like idea that determines its content. Perhaps the analogy to automatic vending machines may make this more perspicuous. Consider two different coin-slots A and B within an automatic vending machine.¹ Coin-slot A is activated by insertion of coin C_A ; coin-slot B is activated by insertion of coin C_B . Note that the answer to the question whether some particular coin-slot is activated by C_A or by C_B is determined by the internal mechanism of the automatic vending machine. In the situation assumed here, the content of coin-slot A (or rather: of coin-slot A's being activated) is that coin C_A is inserted or present, while the content of coin-slot B (or, again, rather: of coin-slot B's being activated) is that coin C_B is inserted or present. However, in the reversed situation, where coin-slot A is activated by insertion of coin C_B and coin-slot B is activated by insertion of coin C_A , the content of the coin-slots (or rather: of the coin-slots being activated) would be reversed too. This shows that the activation of the specific coin-slot A within an automatic vending machine is a signal to the machine that some particular coin has been inserted. Analogously, the activation of the slot-like idea TREE within a subject is a signal to that subject that some particular type of situation in reality has been perceived.

To this line of reasoning several objections can be raised. The first objection concerns the following. Above it was claimed that a subject may determine the type of situation in reality perceived by it via the activation of its ideas, but how can the subject *know* that the activation of a specific idea corresponds to a particular type of situation and not to some other type? A subject that cannot be certain about this

¹ Although there are two coin-slots, the vending machine has only one coin insertion point. The term 'coin-slot' refers to the slots in which coins eventually land, after they have been inserted in the automatic vending machine.

cannot rely upon its ideas, which would render them useless. This problem was also considered by Jerry Fodor, who offered the following explanation:

Notice [...] that the use of a language for computation does not require that one should be able to *determine* that its terms are consistently employed; it requires only that they should in fact *be* consistently employed. Someone might argue like this: 'Imagine a man doing sums, and suppose that he has no way of assuring himself that the number that he was using the numeral '2' to refer to five minutes ago was the same number that he is using the numeral '2' to refer to now. Then surely, he *couldn't* use the numerals to effect his computations'. But, surely, he could. The soundness of inferences is not impugned by the *possibility* of equivocating, but only by instances of actual equivocation. (Fodor 1975, p. 70)

So, in some sense the subject is dependent upon the activation of its ideas for knowing in which situation in reality it finds itself, and therefore the definition of knowing depends on this activation. Hence it is not appropriate to speak of a subject that cannot know and the only usefulness condition for the activation of ideas in a subject is that, as Fodor explains, this activation proceeds in a consistent fashion.

Another objection that can be raised against the concept of ideas having content in virtue of their activation is also considered by Fodor (Margolis and Laurence 1999, Chapter 22), who distinguishes two different ways in which ideas can become activated: labeling and representing. Labeling has been discussed repeatedly above. It occurs when a perceived situation in reality is such that it triggers the internal structure of a subject to start the activation process of some specific idea. Representing has not yet been considered up to this point. It is the capacity of human subjects that is better known as imagining or, somewhat more generally, thinking of. In a situation of representing a specific idea is not activated by some specific type of situation in reality (as in the case of labeling). Instead, the idea is activated by some internal process of the subject. Representing can be compared with a situation where the coin-slots of an automatic vending machine are not activated by the insertion of a specific coin but by some other means, say, by some internal coin-counting process. Since in such situations the slot-like idea is not activated by some situation in reality, it is unclear what this means for the content of such an idea. If content is always determined by the source of activation, then being activated by something else than the type of situation in reality that serves as the usual source of activation, means having a different content. But this does not seem plausible. A possible solution to

this problem would be to say that although at some low level of idea-activation, subjects indeed cannot tell the difference between activation by labeling and activation by representing, human subjects are able to do so at a higher level. For in cases of labeling their senses are stimulated, whereas this does not hold for cases of representing. And it is the labeling cases that determine content, because after all the content of the idea TREE consists of the tree-containing situations in reality by which that idea is activated; not of the thoughts that also happen to activate it.

Although the activation and the development process of ideas were presented as two separate processes, the river analogy suggests that the distinction between them may not be that clear-cut. It can be observed that water flowing through a river also minimally influences the course of the river – and the activation process of ideas may very well do the same thing in the brain.

Chapter 4 Abstract Ideas

This chapter takes up the third sub-question of the present master's thesis, viz., is it possible to meet the abstract-ideas objection against empiricist mentalist semantic theory within an interdisciplinary approach, and if so: how? The chapter consists of two sections. The first one gives an elaborate description of the abstract-ideas objection against empiricist mentalist semantic theory. Several versions of this objection will be treated and confronted with Locke's thoughts about abstract ideas. Moreover, two problems will be argued to lie at the heart of the objection, in the sense that incorporating a solution to these problems would make the resulting theory less vulnerable to the self-applicability objection. Some existing proposals for solving these problems will be discussed and evaluated.

The second section addresses the two central problems in the context of the existing proposals for solving them, proposes some further solutions, and investigates the consequences of these solutions for empiricist mentalist semantic theory in general.

4.1 The Abstract-Ideas Objection

In the first chapter we already shortly described the abstract-ideas objection against empiricist mentalist semantic theory, which will be discussed more thoroughly below. The present section consists of five parts. The first part explores the notion of abstractness. The second part explains the abstract-ideas objection and introduces two derived problems that seem fundamental to it. The third part relates the objection to the work of the prominent empiricist John Locke. The fourth and the fifth part present and evaluate proposals for solving (one of) the derived problems.

4.1.1 Abstract Entities

In this section we will try to get a better understanding the notion of abstractness. What does it mean for something to be abstract?¹ According to O'Connor, abstractness is the capacity 'of representing more individuals than one' (O'Connor

¹ Sometimes the term 'general' is used instead of 'abstract'.

1967, p. 137). This implies that abstractness arises in special cases of representing; viz., cases where more than one individual is represented. Although O'Connor's interpretation of abstractness seems natural, the notion still needs some further clarification. In a situation of representation there are always two entities involved: the entity that represents and the entity that is represented (Lowe 1995, p.p. 159-160). To which of these two entities should the notion of abstractness be applied? Within this section we will use an interpretation of the notion of abstractness according to which it is the entity that is represented that may have the property of being abstract, and not the entity that represents. Now, if abstractness is a property of the entity that is represented, the question arises: which property? Here we may incorporate O'Connor's interpretation and say that abstractness means that the entity that is represented actually comprises multiple individuals. So if the entity that is represented comprises multiple dogs – and this holds for the idea DOG, which represents all dogs, then this entity is abstract because multiple dogs do not exist as a real individual. This interpretation, however, does not include all entities that we tend to call abstract; sometimes the entity that is represented is even less directly related to the world of the senses. This holds, for example, for the idea DEMOCRACY. Here the entity that is represented is abstract but not directly related to sensation. In the rest of this section an entity that is represented will be called abstract if and only if it comprises multiple individuals in the sensation of a subject, such as the multiple entity represented by the idea DOG, or if it is not directly related to the sensation of a subject, such as the abstract entity represented by the idea DEMOCRACY. The first type of abstract entity will be characterized as 'directly related to sensation', while the second type will be characterized as 'not directly related to sensation'. Although the term 'abstract entity' may suggest the existence of some Platonic third realm, this is not the intention of the present master's thesis. The term is merely used because it has some explanatory advantages.

The previous paragraph interpreted abstractness as the property of an (abstract) entity, but often another interpretation is offered. Sometimes an 'abstract entity' is also seen as the result of a specific process, which is called 'abstraction'. For example, Nicholas Jolley notes that "to speak of abstract ideas, for Locke, is thus to say that they result from a certain mental process" (Jolley 1999, p. 50); a hundred pages further he defines this process as "eliminating from a number of particular ideas

everything except that which is common to all” (Jolley 1999, p. 148). It should be observed that Jolley speaks of ideas as if they are themselves abstract entities. This is not the way in which this master’s thesis uses the concept, but let us for now accept Jolley’s terminology. (The next paragraph will supply some further clarification.) The process described by Jolley implies that the subject already has ‘a number of particular ideas’ and these particular abstract entities can be said to be less abstract than the abstract entity that is constructed from them by eliminating everything except ‘that which is common to all’. These less abstract entities of course also develop as a result of this mental process and are themselves based on other, again less abstract, entities. Eventually, this reversal of the abstraction process will, according to empiricists, reach entities that are not abstract anymore but may be directly observed to perception. Although merely invoking a mental process is not enough to explain an (abstract) entity itself, it does raise an important issue with respect the development of abstract entities to which we will return below.

Now that the concept of abstract entities has been made somewhat clearer, the next section will explain and analyze the abstract-ideas objection. The very phrase ‘abstract ideas’ may be confusing in relation to our interpretation of the term ‘abstract’, in view of the fact that abstractness has just been defined as a property of the entity that is represented, and not as a property of the entity that represents. The phrase ‘abstract ideas’ suggests that abstractness is a property of idea, so that it is the idea that is abstract, but this is not the correct interpretation. The phrase ‘abstract idea’ should be read as referring to a situation of representation in which the idea is the entity that represents and in which the entity that is represented has the property of being abstract (in the interpretation given above). Jonathan Lowe also recognizes this when he says that abstract or general “ideas are not ‘general’ in the sense of being general *things* (universals)” (Lowe 1995, p. 155). Despite this possible confusion, we will keep on using the phrase ‘abstract idea’ for the sake of readability and compatibility with other literature, but what is meant by it is ‘idea that represents an abstract entity’.

4.1.2 The Objection and Two Derived Problems

It has been claimed by some authors that empiricist mentalist semantic theory cannot account for abstract ideas. This is a problem, since human subjects do have abstract

ideas such as DEMOCRACY and because these ideas take part in the semantic phenomenon empiricist mentalist semantic theory should be able to explain it. Thus the abstract-ideas objection states that because empiricist mentalist semantic theory cannot account for the fact that ideas can represent abstract entities the theory must be wrong. The present section discusses some of the objections involving abstract ideas that have been raised against empiricist mentalist semantic theory.

The philosopher George Berkeley, a contemporary of John Locke, argues against abstract ideas along the following lines. About an idea that concerns some part of the human body Berkeley claims that ‘it must have some particular shape and color. Likewise the idea of man that I frame to my self, must be either of a white, or a black, or a tawny, a straight, or a crooked, a tall or a low, or a middle-sized man’ (Bennett 1971, p. 36). Berkeley, who interprets ideas as images (a common practice in the 17th century, see Dunn 1992, p. 118), then argues that indeterminate images do not exist. An image of a human cannot be indeterminate with respect to the color of the human being; it *must* have a color. Therefore, according to Berkeley, ideas do not have the capability of representing abstract entities in the sense described in the previous section. The logical laws that are at the basis of this problem are the law of excluded middle and the law of non-contradiction (Lowe 1995, p. 158). This can be seen as follows. As argued above, an abstract idea is an idea that represents an abstract entity and Berkeley claims that in order for ideas to be able to do this, they must have the capability of being indeterminate. This entails that indeterminacy is a property of the entity that represents, in this case an idea. However, whenever the entity that represents is abstract, i.e., indeterminate with respect to having a potential property, such as some particular color, it violates the above logical laws. For note that according to the law of excluded middle the entity should either have or not have the property and according to the law of non-contradiction the entity cannot both have and not have the property. Thus, since no entity can violate the logical laws of excluded middle and the law of non-contradiction, no entity can be indeterminate. That is to say: no entity can represent an abstract entity and empiricist mentalist semantic theory is incorrect.

Martin Stokhof (2003, pp. 33-34) offers another variant of the abstract-ideas objection against empiricist mentalist semantic theory. He writes: “Good examples are (names for) abstracta such as: ‘four-dimensional space-time’, ‘democracy’, ‘God’,

'free will'. How are these ideas obtained from sense impressions through abstraction?"¹ This objection against abstract ideas is a specific version of a more general objection against empiricist mentalist semantic theory. Note that above an approach to abstractness was sketched according to which an abstract entity is the result of an abstraction process. This abstraction process reflects the central claim of empiricism that sensation is the source of all ideas. Now, the more general objection states that for most ideas (not only abstract ones) it is unclear how they derive from sensation. Jerry Fodor et al. argue along these lines in the below passage. They focus on definitions, but their objection applies to ideas as well:

If there are few convincing examples of definitions, there are literally *no* convincing examples of definitions which take *prima facie* non-sensory/motor terms into a sensory/motor vocabulary. [...] the 'right hand' (definiens) side of standard examples of definitions does not tend, by and large, to be more sensory/motor than the 'left hand' (definendum) side. So for example the conventional wisdom has it that 'grandmother' means 'mother of father or mother'; 'ketch' means 'two masted sailing vessel such that the aft (or 'mizzen') mast is stepped forward of the helm'; 'bachelor' means 'unmarried man'; 'kill' means 'cause to die,' etc. It's hard to discern a tendency toward the sensory/motor in these examples, and these examples are quite typical. Surely, no one could seriously argue that words like 'mother, father, mast, vessel, cause, die, married and man,' all of which occur in the definiens, are somehow closer to transducer language, than say, 'grandmother,' 'ketch,' 'kill' or 'bachelor' are. (Margolis and Laurence 1999, p. 494).

Thus, Stokhof and Fodor et al. point to some problematic aspects of abstract ideas. One problem concerns the question how abstract ideas are related to sensation. It is clear that some abstract ideas are not directly related to sensation, although, at least according to empiricists, they are somehow derived from sensation. How is this gap bridged? Fodor et al. note that this gap is usually assumed to be bridged by compositionality. That is, more complex ideas are supposedly built up from less complex ones, and in last analysis from ideas that are directly related to sensation. But Fodor et al. also show that it is not at all clear whether compositionality is up to this task. Another problem concerns the question how the development process of abstract

¹ My translation of: "Goede voorbeelden zijn (namen voor) abstracta, zoals: 'vierdimensionale tijd-ruimte', 'democratie', 'God', 'vrije wil'. Hoe zijn deze ideeën via abstractie te verkrijgen uit zintuiglijke impressies?" (Stokhof 2003, pp. 33-34).

ideas works: it is not at all clear how abstract ideas that are not directly related to sensation are constructed from other ideas.

A last objection is based on the indistinctness of the account of abstract ideas that is offered by empiricist mentalist semantic theory. In particular, Nicholas Jolley criticizes the indistinctness of the development process of abstract ideas: ‘Unfortunately, however, it is less than clear whether he [John Locke] offers a single consistent account of this mental process whereby abstract ideas are formed’ (Jolley 1999, p. 50). Jonathan Lowe also refers to this process (Lowe 1995, p. 155) as one “which Locke attempts to illustrate by examples but never succeeds in defining at all exactly.”

The above paragraphs illustrate the abstract-ideas objection against empiricist mentalist semantic theory. The gist of the objection is that a subject has an abstract idea whenever it has an idea that represents either something that comprises more than one individual or something that does not exist in sensation, but that opponents claim that the ideas that figure in empiricist mentalist semantic theory are incapable of doing this.

From the various above objections two problems can be derived that should be solved by empiricist mentalist semantic theory.

The first problem concerns the question how ideas can represent abstract entities. The issue of the nature of ideas in general was already addressed in the chapter about the self-applicability objection and will not be taken up again, but it is also very relevant to the problems discussed in the present chapter. The previous chapter explored several possible interpretations of ideas such as images and objects but it is not at all obvious whether (and, if so, how) these various interpretations are capable of representing abstract entities.

The second problem concerns the question how (abstract) ideas actually develop. This problem involves the explanation of abstractness as the result of undergoing the process of abstraction that was discussed earlier in the present chapter and that was called the development process of ideas in previous chapters. The importance of these two problems abstract ideas are also acknowledged by Jonathan Lowe when he says that “his [John Locke’s] next two questions must be: what sort of ideas do general terms signify, and how do we acquire those ideas?” (1995, p. 155).

4.1.3 The Objection Applied to Locke

The previous subsection discussed some variants of the abstract-ideas objection against empiricist mentalist semantic theories. In the introductory chapter it was announced that Locke's *An Essay Concerning Human Understanding* will be used as the main representative of empiricist mentalist semantic theory, and Locke has in fact already been mentioned a couple of times above. This subsection will further explore the relation between the variants of the objection and Locke's empiricist mentalist semantic theory and show that the objections are indeed valid objections against it.

The first argument was that of George Berkeley, who argued that Locke's abstract ideas are like images, so that since images cannot be indeterminate, ideas cannot be indeterminate either. Now, according to Locke, indeterminacy is indeed necessary for abstractness, as becomes clear from a somewhat paradoxical (cf. Ayers 1991, pp. 57-58) passage in *Essay Concerning Human Understanding*:

Does it not require some pains and skill to form the general Idea of a Triangle [...], for it must be neither Oblique, nor Rectangle, neither Equilateral, Equicrural, nor Scalennon; but all and none of these at once (Locke 1690, Section IV.vii.9).

If something is neither A nor B it can be said to be indeterminate with respect to A and B, so Locke does claim here that ideas are indeterminate, just as Berkeley said. But does Locke also claim that ideas are like images? He does not; but he does not claim the contrary either. Locke is quite vague about the nature of ideas. For Locke, an idea is "the Object of the Understanding when a Man thinks, I have used it to express [...] whatever it is, which the Mind can be employ'd about in thinking" (1690, Section 1.1.8). But there are also passages where ideas are described in image-like terms. Thus: "To ask, *at what time a Man has first any Ideas*, is to ask, when he begins to perceive; having Ideas, and Perception being the same thing" (Locke 1690, Section II.i.9), and: "[t]he *Ideas* of the Nurse, and the Mother, are well framed in [Children their] Minds; and, like Pictures of them there, represent only those Individuals" (Locke 1690, Section III.iii.7). Locke's vagueness as regards ideas gave Berkeley the chance to interpret the concept in a way that enabled him to raise his valid objection.

The objection by Stokhof stated that it is not at all obvious how abstract ideas such as FREE WILL or DEMOCRACY are derived from sensation using abstraction. Now,

Locke does give some examples of how some ideas are abstracted from our sensations. About the abstraction process that leads to the idea MAN within children, Locke says “they [the children] make nothing new, but only leave out of the complex Idea they had of Peter and James, Mary and Jane, that which is peculiar to each, and retain only what is common to them all” (1690, Section 3.3.7). Elsewhere, Locke describes how this process continues to form the abstract idea ANIMAL, “which new Idea is made, not by any new addition, but only as before, by leaving out the shape, and some other Properties signified by the name Man, and retaining only a Body, with Life, Sense, and spontaneous Motion, comprehended under the Name Animal” (Locke 1690, Section III.iii.8). However, although Locke’s description of the abstraction process of ideas explains the creation of *some* abstract ideas it still remains unclear how the process may result in ideas such as FREE WILL and DEMOCRACY.

4.1.4 Some Proposals

We have seen that there are two unanswered questions with respect to abstract ideas. Viz.: how can ideas represent abstract entities? And: how do ideas that represent abstract entities develop? This section discusses some answers to these questions that can be found in the literature.

Jonathan Lowe describes a solution to the first problem. In the following passage he suggests an interpretation of Locke as regards the nature of abstract ideas:

The suggestion is thus that the process of abstraction consists in our comparing our ideas of various particulars which we encounter in sensation, noting their similarities and differences, ignoring the latter and retaining in mind only the former (the similarities) as a sort of pattern or template, which we may then go on to employ in classifying further particulars that we meet: and these patterns or templates are our *abstract general ideas*. (Lowe 1995, p. 156)

So on Lowe’s view of Locke, abstract ideas can be seen as *patterns*, i.e., similarities in sensations that instantiate the abstract idea. According to this reading, abstract ideas have a pattern-like nature. The traditional view explains the nature of abstract ideas in terms of objects such as images, stereotypes or mental entities but all these are subject to objections along the lines of Berkeley’s. Lowe’s alternative approach of analyzing ideas as patterns should be capable of representing the two types of abstract entities that were discussed above.

Also Locke himself hinted at a pattern-like view of abstract ideas, as becomes clear from the following two passages:

[T]he *Idea* annexed to [a] Name would not cease to be what it is; nor cease to be as a pattern, to determine which of the particular Figures we meet with, have, or have not a Right to [that] Name. (Locke 1690, Section III.iii.19)

The *Names of simple Ideas and Substances*, with the abstract *Ideas* in the Mind, which they immediately signify, [...] from which was derived their original pattern. (Locke 1690, Section III.iv.2)

Neither Locke nor Lowe (nor Michael Ayers 1991, see pp. 248-249) further elaborate on the hypothesis that ideas are pattern-like, but it may be worthwhile to go further into its possibilities and potentials. Pattern-like ideas may have specific properties that may enable them to avoid some of the problems described above.

The second problem involves the question how ideas that represent abstract entities develop, or in other words: what does the development process of (abstract) ideas look like. Unfortunately, not much progress has been made with respect to this problem since John Locke and there are no real proposals that explain how abstract ideas such as TRIANGLE and DEMOCRACY may develop. Lowe (1995, p. 157) does mention connectionist models of pattern recognition in his description of abstract ideas. Since his proposal for solving the first problem suggests that ideas are patterns, such models of pattern recognition and especially the techniques used to develop them may make an interesting contribution to solving the second problem.

4.1.5 Do the Proposals Work?

In this section we will shortly evaluate Jonathan Lowe's proposal for solving the two problems concerning abstract ideas that was discussed in the previous section and describe possible shortcomings and objections.

One possible objection is that an explanation of the nature of abstract ideas as pattern-like entities is still quite vague: it is unclear how the patterns are realized and in what exact sense these patterns are responsible for the abstract ideas. Furthermore, while the first type of the two types of abstract entities that were introduced above might be explainable in terms of patterns, it is not at all clear how this could be made to work for the second type. The abstract entities involved here are not directly related

to the realm of sensations and some account should be given of how this type of entities may succeed in escaping this domain.

Next, even assuming that the first problem can be solved so that it is possible for pattern-like ideas to represent abstract entities of the two types that were introduced above, we still need an account of how these pattern-like ideas may come to exist. Jonathan Lowe appeals to computer science and connectionism for a possible answer to the first problem and this might also be the direction in which to look for answers to the second problem of how these pattern-like ideas develop. Again, this ‘solution’ is quite vague and more details are required in order to be able to properly discuss it.

In this section we have given an overview of the abstract-ideas objection against empiricist mentalist semantic theory. In doing so, we have discussed the term ‘abstract’, clarified the abstract-ideas objection using the work of, among others, Berkeley and Fodor, and sketched existing proposals for solving it. We have also argued that these proposals still require further elaboration, and this is what we will focus on in the next section.

4.2 Meeting the Abstract-Ideas Objection

This section focuses on the way in which empiricist mentalist semantic theory may meet the abstract-ideas objection. A detailed description of this objection was given above, where the objection was partitioned into two underlying problems that need further clarification and some proposals from the literature were discussed. The present section goes further into these problems and their solutions, using the proposals from the literature as a guideline. The problem concerning the nature of abstract ideas will be handled in its first part, while the second part will discuss the problem regarding the development process of abstract ideas.

4.2.1 The First Problem

As explained above, the term ‘abstract idea’ can be misleading because abstractness does not refer to a property of ideas, but rather denotes a property of the entities that are represented by them. Therefore, this section will go into the question how ideas represent abstract entities. We already discussed Berkeley’s objection, who argued that Locke must require that ideas can be indeterminate in order for them to be

able to represent abstract entities, a requirement which Berkeley considered an absurdity for the image-like objects he took ideas to be. Of course, Berkeley's objection makes sense for image-like ideas. It can only be met by adopting an alternative approach to ideas that does provide them with the capability of being indeterminate. We have also seen that Lowe suggested that ideas could be pattern-like, but his account lacked sufficient detail for showing how pattern-like ideas may actually represent abstract entities. The present section will try to add the required detail by going further into patterns and use them in an account of how ideas may represent abstract entities. Above two types of abstract entities were distinguished, and these will be treated separately below. Recall that the first type of abstract entities is directly related to sensation, while the second type is not directly related to sensation. However, before showing how the two types of abstract entities can be handled using patterns, some important issues need to be addressed.

First, it is important to distinguish between the two sources of ideas that were discerned by Locke, viz., sensation and reflection. Sensation comes in via the sensors of the subject that receive their information from the outside world. Reflection takes place in the internals of the subject, which are themselves a source for ideas such as THINKING, PAIN and BELIEVING. Why this distinction is important will become clear soon.

Second, it is important to reflect on what patterns exactly are. Because patterns will play a fundamental role below, some terminology and theory will come in handy. A *pattern* may be defined as a regularity in a *string*¹ written in a specific *language*. By way of illustration, assume that the language is built up from letters of the alphabet combined with the ‘_’ character. A string, then, is simply a series of elements from the language such as “__boooooom__” or “boom__”. Now, a pattern that occurs in both strings is: it begins with one **b**, followed by two or more **o**'s, followed by one **m**, with which it also ends!² Otherwise put: the two strings conform to that pattern. A frequent misconception is that a pattern is something which can be described within the language itself, but this is not true. Note, for instance, that the language of the

¹ In computer science regularities in strings can be expressed by Regular Expressions. There is a lot of literature on this subject. See for example (Friedl 1997).

² Strings will be enclosed by “ ”, while patterns will be surrounded by | |.

above example is not equipped with predicates such as ‘begin with’, ‘followed by’, ‘two or more’ and ‘end with’, which are *about* the language. For clarifying the often complex nature of a pattern a *pattern recognizer* may be invoked. A pattern recognizer is the (physical) realization of a pattern in strings of some specific language. It is a device which takes as its input a string written in the language and has an output that becomes active when the string conforms to the pattern and remains inactive when the string fails to do so. Although this is perhaps still not completely clear, further discussion of pattern recognizers will be postponed until the second half of the present section.

We will now discuss how patterns can account for ideas that represent the first type of abstract entities. Berkeley’s objection claims that representing such abstract entities requires that ideas be indeterminate. He argued that image-like ideas must fail in this respect, because images cannot be indeterminate. The goal of the present section is to show that pattern-like ideas may succeed here, because they can be indeterminate. This demonstration will proceed in two steps. The first step consists of showing that pattern recognizers can be indeterminate. The second step transfers these results to pattern-recognizing ideas, showing that pattern-recognizing ideas can be indeterminate as well.

Let us first show, then, that pattern recognizers can be indeterminate. Indeterminacy, as already explained in this chapter, can be analyzed as the violation of at least one of two fundamental logical laws: the law of excluded middle and the law of non-contradiction. That is to say: an indeterminate idea represents an abstract entity, and this means that the entity that is represented violates excluded middle and/or non-contradiction. Observe that indeterminacy is a property of the entity that represents and that violation of the logical laws is a property of the entity that is represented. The first logical law involved is the law of excluded middle. This law states that for every entity and every property the entity either has the property or fails to have it; there is no middle between the two. The following example may serve to illustrate the fact that it is possible for a pattern recognizer to represent an entity that neither has a certain property nor fails to have that property. Imagine a pattern recognizer that recognizes the following pattern: begins with **bo**, followed by either one **k** or one **o**, followed by **om**, with which it also ends. Now consider the property of having a central character **k**. It is obvious that the abstract entity that is represented

by the pattern recognizer neither has nor fails to have the property of having a central character **k** and therefore violates the law of excluded middle in some sense. The second logical law is the law of non-contradiction. This law states that for every entity and every property the entity cannot both have and not have that property. If the above pattern recognizer and property are again considered, it may be noted that the law of non-contradiction is also violated in some sense, for the abstract entity represented by the pattern recognizer does possess the property of having a central character **k** but is also not in possession of that property. So, pattern recognizers are capable of representing entities that somehow violate both logical laws and therefore have the capability of being indeterminate.

The next step involves the translation of the above in terms of pattern-like ideas. This requires determining the correspondences between the theory of pattern recognition and empiricist mentalist semantic theory. A straightforward suggestion is to let pattern-like ideas correspond to pattern recognizers, so that ideas will also realize a pattern within a string written in some language. The language in which the strings are written depends on the type of subject involved. In the case of human subjects the strings will probably be concatenations of electrical and chemical signals within its brain and in the case of computer subjects the strings will probably be concatenations of electrical currents. Notwithstanding the difference in language, the pattern-recognizing ideas will have an output that also becomes active whenever an input string conforms to the pattern realized by the idea and does not activate otherwise and remains inactive when the input string fails to do so. Recall the example of John Locke that was discussed earlier:

Does it not require some pains and skill to form the general Idea of a Triangle [...], for it must be neither Oblique, nor Rectangle, neither Equilateral, Equicrural, nor Scalenon; but all and none of these at once. (Locke 1690, Section IV.vii.9)

If the translation from pattern recognizers to ideas is viable, the indeterminacy that Locke describes in this passage should apply to the pattern-recognizing idea TRIANGLE as well. Closer consideration of this particular pattern-recognizing idea reveals that the concepts of ‘language’ and ‘string’ that were introduced above will

obviously receive a different interpretation. The ‘language’ does no longer consist of the alphabet, but comprises the signals that the subject receives from its sense organs. The interpretation of the concept ‘string’ will also be different. Imagine that the signal from a certain sense organ of the subject is sliced into multiple snapshots. Then a string will be an input signal from a subject’s sense organ at a specific moment in time. Let us focus on the visual sense organ of a subject in order to elucidate the pattern-recognizing idea TRIANGLE. Comparable to – but much more complex than – the linguistic pattern recognizers introduced above, it can be imagined (with some effort) that the pattern-recognizing idea TRIANGLE is activated by all and only strings of input signals that stem from visual experiences of triangular figures. Accordingly, the entity represented by the pattern-recognizing idea would be “neither Oblique nor Rectangle, neither Equilateral, Equicrural, nor Scalenon; but all and none of these at once”, just as the entity represented by the pattern recognizer *l* begins with one **b**, followed by either one **k** or one **o**, followed by **om**, with which it also ends; neither has nor lacks a central character **k**, “but [both] and none of these at once”.

This means that Locke’s passage is not only not paradoxical, but that entities represented by pattern-recognizing ideas are also somehow able to violate the logical laws of excluded middle and non-contradiction and therefore pattern-recognizing ideas can be indeterminate as well. Because indeterminacy is required to represent the first type of abstract entity, this also lends some plausibility to the suggestion that ideas with pattern-recognizing properties may represent the first type of abstract entities and thus meet Berkeley’s objection that ideas must fail to represent abstract entities because they cannot be indeterminate.

Recall that we also distinguished another type of abstract entities, viz., those that are not directly related to sensations. An example of the second type of abstract entity is the one represented by the idea DEMOCRACY. Accounting for such abstract entities along the lines sketched above does not work, because the pattern recognizers, though indeterminate, are directly related to their input. They can only accept or reject strings on the basis of the presence of specific elements of the language at certain positions in the string. In order to be able to handle the second type of abstract entities, pattern recognizers will need an additional source of input that will allow them to exploit additional information within the patterns they recognize. This additional source of input concerns information on whether other pattern recognizers

are activated by the same input string. As a result, input strings will not only have to realize the pattern of some particular pattern recognizer, but also have to meet the conditions imposed by – i.e, to realize the patterns of – other pattern recognizers present in the subject. We will now proceed to address two issues raised by this additional source of input. First, we will show that it may indeed result in a broader range of pattern recognizers. And second, we will translate the resulting situation in terms of pattern-recognizing ideas.

So, let the language again be built up from of the letters of the alphabet and let the pattern recognizer TREE be as follows: lbegins with one **b**, followed by two or more **o**'s, followed by **m**, with which it also endsl. Strings that activate the pattern recognizer TREE are “boom”, “boooooom” and “booom”. Strings that fail to do so are “bom”, “bioom” and “bm”. Next, consider the three pattern recognizers ROOT, TRUNK and TOP, where ROOT is lpattern recognizer TREE activates, one **bl**, TRUNK is lpattern recognizer TREE activates, two or more **o**'sl and TOP is lpattern recognizer TREE activates, one **ml**. Note that these pattern recognizers rely upon the pattern recognizer TREE, so that strings such as “bim” fail to activate ROOT and TOP. Observe that the following two pattern recognizers are in turn related to the TRUNK pattern recognizer: the pattern recognizer SHORT is lpattern recognizer TRUNK activates, two, three or four **o**'sl and the pattern recognizer LONG is lpattern recognizer TRUNK activates, at least five or more **o**'sl. The pattern recognizer SHORT is only activated by trees with short trunks such as “booom”; the pattern recognizer LONG is only activated by trees with long trunks such as “boooooom”. Finally, consider the pattern recognizer HEIGHT: leither pattern recognizer SHORT or pattern recognizer LONG activatesl. This pattern recognizer is reminiscent of the abstract entities that do not have a direct relation to sensation: trying to pinpoint some part of a string that directly relates to the pattern recognizer HEIGHT is just as difficult as pinpointing some part of sensation that directly relates to the idea DEMOCRACY. Thus the pattern recognizer HEIGHT represents a type of abstract entity that has no direct relation to sensation.

The next challenge is to translate the situation that results from introducing the additional source of input in terms of pattern-recognizing ideas. Obviously, what is required is a source from which information about the activation of other pattern-recognizing ideas is made available to the pattern-recognizing idea under

consideration. Here we may find support in Locke's distinction between two sources of ideas that was alluded to above and can be detected in the following passage:

Our Observations employ'd either about *external, sensible objects; or about the internal Operations of our Minds, perceived and reflected on by our selves, is that, which supplies our Understandings with all the materials of thinking.* These two are the Fountains of Knowledge, from whence all the *Ideas* we have, or can naturally have, do spring. (Locke 1690, Section II.i.2)

Locke calls the two sources sensation and reflection. When these are compared with pattern recognizers the parallelism is obvious. What Locke means by sensation is the pattern-recognizing-idea counterpart of the string input ('external sensible objects') of all pattern recognizers in a subject. And what Locke means by reflection is the pattern-recognizing-idea counterpart of the additional input consisting of information on whether other pattern recognizers within the subject ('the internal operations of our minds') are activated by the same input string. Note that a subject may contain many pattern-recognizing ideas and each of these pattern-recognizing ideas is based on the sensation and/or reflection of that subject. Sensation connects the subject to the outside world. Reflection operates within the subject and is based on access to information about the activation of all other pattern-recognizing ideas. By way of illustration, consider the following situation in which reflection and sensation are combined within pattern-recognizing ideas.

As above, we shift from a language built up from the letters of the alphabet to a language that concatenates input signals from the visual sense organ of a subject. Thus, the input signals caused by a tree activate some pattern-recognizing idea TREE. In addition to this, the three pattern-recognizing ideas ROOT, TRUNK and TOP are derived in a similar way as sketched above and the same holds for the pattern-recognizing ideas LONG and SHORT. Now consider the pattern-recognizing idea HEIGHT: it is described as either pattern-recognizing idea SHORT or pattern-recognizing idea LONG activates. Answering the question what it is that the entity represented by the pattern-recognizing idea HEIGHT is related to in sensation, or what sensations cause the pattern-recognizing idea HEIGHT to become active, is quite difficult because the idea does not directly represent sensation: just as the abstract entity that is represented by the idea DEMOCRACY, the entity that is represented by the pattern-recognizing idea HEIGHT has the property of not being directly related to sensation.

A fully satisfactory reply to Fodor's abstract-ideas objection requires an explanation of how ideas such as GRANDMOTHER, KETCH, KILL and BACHELOR depend on the activation of other ideas. However, there are three reasons why this is difficult. The first reason is complexity. In order to illustrate how pattern-recognizing ideas can represent abstract entities that are not directly related to sensation, we have employed a relatively simple idea: HEIGHT. However, this does not mean that all pattern-recognizing ideas that represent abstract entities can be as easily explained in terms of other pattern-recognizing ideas; it is more likely that ideas such as DEMOCRACY (or the ones mentioned by Fodor) are far more complex. The second reason is ambiguity. In our example there exists only one explanation in terms of pattern-recognizing ideas for the idea HEIGHT, but more probably, in view of the fact that most pattern-recognizing ideas within human subjects are actually far more complex, each human subject has a slightly different version of particular pattern-recognizing ideas. Of course, this considerably complicates the task of accounting for specific pattern-recognizing ideas, because there is not one single account! A third reason is vagueness. We have described pattern-recognizing ideas as entities that either become active or remain inactive on the basis of a certain input, but it may be more plausible (at least in the case of human subjects) to describe them as less discrete entities, in that pattern-recognizing ideas become active to a certain degree on the basis of a certain input.

Abstract entities have been discussed extensively above, but their ontological status has not been discussed. It was already noted that this discussion primarily served explanatory purposes and had no Platonic intentions. Abstract entities do indeed have a mainly explanatory role to play within the theory of pattern recognizers, for a pattern recognizer represents an abstract entity because it is activated by certain patterns in the input string. Although the abstract entity that is represented has no real existence within the input string, the activation pattern of the pattern recognizer does raise the suggestion of its existence.

What has been argued in this section is that given pattern-recognizing ideas, the two types of abstract entities under consideration are considerably more plausible than with image-like interpretations of the nature of ideas. Just as pattern recognizers, pattern-recognizing ideas represent abstract entities because they are activated by certain input strings, an activation pattern that is realized by the internals of the

subject. The next section discusses how ideas that represent abstract entities are formed in an attempt at providing an account of how the internals of a subject develop in such a way that the subject's pattern-recognizing ideas actually become activated by certain input patterns.

4.2.2 The Second Problem

The second issue that requires clarification is the question how abstract ideas are formed. We saw above that according to Locke, the process is subject to the constraint that every abstract idea is eventually derived from sensation. But this development process is far from obvious. The previous section argued that in addition to the sensation of a subject also its reflection is required for having ideas represent all types of abstract entities. This complicates the development process of ideas and therefore we will initially only take sensation into account. After that we continue with an account of the more complicated development process of ideas that incorporates both sensation and reflection.

An analysis of the development of something may benefit from a closer look at what is being developed. In the previous section abstract ideas were analyzed as pattern-recognizing devices that are capable of representing abstract entities because they are activated by certain input patterns. Such ideas constitute a special type of pattern recognizers and therefore we go further into the nature of the latter. Pattern recognizers are actually quite common in everyday life. An example of a simple pattern recognizer is a motion detector as used, for example, in outdoor lamps. These lamps start to burn whenever they face an input string that conforms to the pattern recognizer for motion that is built into it. In the case of Passive Infrared Sensors the input string consists of several snapshots at successive moments in time. The string conforms to the pattern MOTION whenever there is a rapid change between the different snapshots in signals caused by infrared light of a certain wavelength. Speech recognition systems constitute another example of a pattern recognizer. These systems convert a speech signal to a sequence of words. The input strings for the system consist of input signals from a microphone and computer software simulates several pattern recognizers, where each recognizes a specific word. Strings of speech signals are fed to the pattern recognizers simulated by the computer software, which tell whether their pattern is present in the string. From this information a sequence of

words is built that is supposed to correspond with the speech signal from the microphone. There are numerous more examples: security cameras with microphones that detect violent behaviour, automatic license number recognizers that guard the speed limit in traffic, fully automated fruit sorting systems that select fresh fruit for distribution to the consumer, medical software that automatically detects abnormalities in scans of entrails – the applications of pattern recognizers become wider every year.

All these examples embody the principle that pattern recognizers are devices that are structured in such a way that whenever they are given a certain input, they produce an output that decides whether the input conforms to the pattern recognized by the device. How exactly these devices are structured depends on the particular implementation of the pattern recognizer. It may be implemented using the binary structures of a computer (like speech recognition systems) or, alternatively, using cogs, levers and springs. In practice, the actual structure of the device is just the optimal way to realize the pattern recognizer in a particular context, but the pattern recognized does not depend on it. In artificial intelligence, this independency of implementation is called multiple realizability. So patterns are multiply realizable in various specific pattern recognizers, and asking how pattern recognizers are formed is asking how their implementation-independent internal structure comes to be such that it is activated by certain strings and not by other ones.

Let us now take a closer look at the development process of pattern recognizers. Even when attention is restricted to computers, which are studied in the field of pattern recognition, there are already several different techniques for implementing a pattern recognizer. A pattern recognizer may, for example, be implemented in a computer using Bayesian decision theory, multilayer neural networks or stochastic methods. Each of these techniques has its own learning algorithm for ensuring that a pattern recognizer implemented using the technique works correctly (i.e., becomes active if and only if the input string conforms to the pattern it is meant to recognize). Most of these learning algorithms rely on humans to determine what counts as a correct or incorrect output for a specific input, but some do not. The latter are gathered under the heading ‘unsupervised learning’ (as opposed to ‘supervised learning’). Going more deeply into unsupervised learning would be beyond the scope of this master’s thesis (and there is already a lot of literature on this subject (see Duda,

Hart and Stork 2001)), but what should be clear by now is that different types of pattern recognizers have very diverse development processes, many of which have actually already been realized within, for example, computer science.

The above can be translated in terms of pattern-recognizing ideas. Within a subject, pattern-recognizing ideas are structured in such a way that they are activated by certain patterns present in input from sensation. The patterns they recognize are also independent of their specific realization. Pattern-recognizing ideas develop within a subject via techniques that probably involve a combination of supervised and unsupervised learning. Thus for at least a decade from the age of four, children undergo an extensive education programme in which their learning is supervised by teachers. But from a very young age, children also have the ability to recognize their mother, the acquisition of which is probably not supervised. Other classification skills, such as the discrimination of different flavours and tastes, are also acquired in an unsupervised way at a very young age. A combination of supervised and unsupervised learning may therefore be expected to account for the development of pattern-recognizing ideas within a subject.

While this may serve as an analysis of the development process of ideas that only involve sensation of a subject, we have seen above that there is another input source on which the activation of pattern-recognizing ideas depends: reflection. This complicating factor of the development process will be first discussed in terms of pattern recognizers and then translated to pattern-recognizing ideas.

Consider, in order to study the complex development process of pattern recognizers, two layers of pattern recognizers. The pattern recognizers in the first layer are only activated by patterns present in the input string, while the pattern recognizers in the second layer are activated by patterns present in the input string combined with information about activation of pattern recognizers in the first layer. The pattern recognizers in the first layer are RED, BLUE, YELLOW, CIRCLE, SQUARE and TRIANGLE. The pattern recognizer RED realizes the pattern `lone r|`; the pattern recognizer BLUE realizes the pattern `lone b|`; the pattern recognizer YELLOW realizes the pattern `lone yl`. The pattern recognizer CIRCLE realizes the pattern `lone character from the alphabet|`; the pattern recognizer SQUARE realizes the pattern `ltwo characters from the alphabet|`; and the pattern recognizer TRIANGLE realizes the pattern `lthree characters from the alphabet|`. The second layer consists of two pattern recognizers:

COLOUR and SHAPE. COLOUR realizes the pattern recognizer RED, BLUE or YELLOW activates and SHAPE realizes the pattern recognizer CIRCLE, SQUARE or TRIANGLE activates. Let us now consider the development process of these pattern recognizers. The pattern recognizers RED, BLUE, YELLOW are respectively activated by the presence of the character **r**, **b** and **y** somewhere in the input string. Learning techniques along the lines sketched above can be used to develop these three pattern recognizers and the same holds for the pattern recognizers CIRCLE, SQUARE and TRIANGLE, which are respectively activated by input strings consisting of one, two and three characters. The pattern recognizers in the second layer make use of information about the activation of the pattern recognizers in the first layer. The second layer pattern recognizer COLOUR is activated by the activation of one the pattern recognizers RED, BLUE and YELLOW in the first layer. This pattern recognizer may develop by making a distinction between the common properties of these pattern recognizers RED, BLUE and YELLOW on the one hand, and the common properties those of pattern recognizers CIRCLE, SQUARE and TRIANGLE on the other. These sets of common properties are distinguishable because the former are activated by the presence of various characters in the input string, while the latter are activated by various lengths of the input string. Of course, the pattern recognizer SHAPE may develop in a completely parallel way.

A similar development process can be envisaged for pattern-recognizing ideas. As before, the language of the input strings consists of concatenations of input signals from the visual sense organ of the subject. The internals of the subject must develop in such a way that its structure implements the above pattern recognizers. At the start of the development process the subject has not developed any structure yet, so that no pattern-recognizing ideas are present. Some combination of supervised and unsupervised learning results in the development of the first pattern-recognizing ideas, say RED, BLUE and YELLOW. These pattern-recognizing ideas are based on the visual sense organ's ability to distinguish between different wavelengths of light. A similar process is responsible for the pattern-recognizing ideas CIRCLE, SQUARE and TRIANGLE. These ideas are based on the visual sense organ's ability to distinguish between different locations from which the light waves enter the sense organ. With the development of these pattern-recognizing ideas a layer has emerged that can be used as additional input for a second layer of pattern-recognizing ideas (see Figure 8).

The development of this second layer requires that the subject have the ability to use reflection as a source of input besides sensation. With similar techniques as the ones outlined above the second-layer pattern-recognizing idea COLOR may develop due to the ability to distinguish between the common properties of the input by which RED, BLUE and YELLOW are activated and the common properties of the input by which CIRCLE, SQUARE and TRIANGLE are activated. These sets of common properties are distinguishable because the former are activated by the presence of light of various wavelengths in the input, while the latter are activated by various locations from which a light wave may enter the input. The same abilities are responsible for the development of the other second-layer pattern-recognizing idea SHAPE.

While, of course, many more details need to be filled in, the above hopefully sheds some light on the development process of ideas that represent abstract entities by its suggestion that this process can be understood as the development of pattern-recognizing ideas as the components of the internal structure of a subject. A wide variety of techniques for creating such structures have been developed in computer science. The application of these techniques to the sensation and reflection of subjects may be expected to contribute to the understanding of the development process of subject-internal structures.

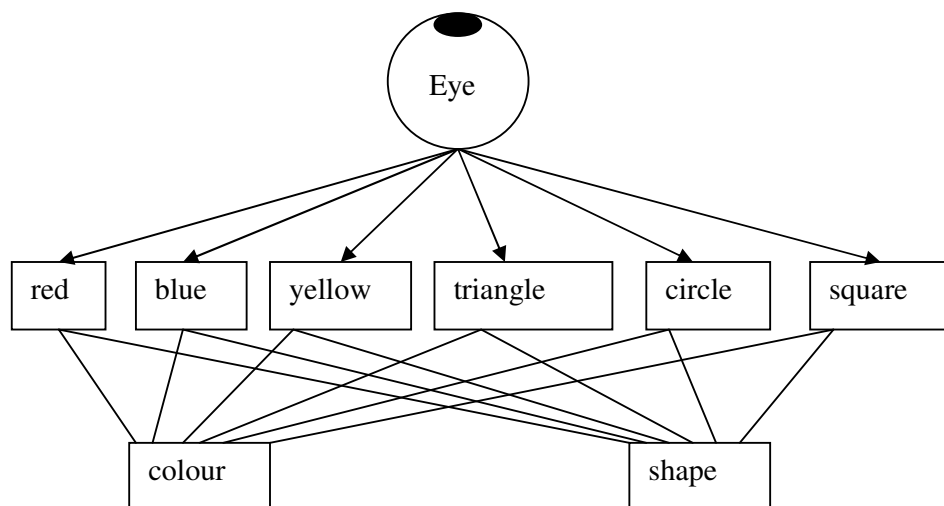


Figure 8

Chapter 5 Conclusion

In this concluding chapter the three sub-questions of the main research question of this master's thesis that were discussed separately in the previous chapters will be put together again in a unified picture of an *updated* empiricist mentalist semantic theory. This unified picture will be sketched in the first section of the present chapter. The second section of this chapter addresses further research questions that arise in the course of this process of unification.

5.1 Findings

The three sub-questions that are discussed in this master's thesis were extracted from its main research question, which reads:

Is it possible to meet the intersubjectivity objection, the self-applicability objection and the abstract-ideas objection against empiricist mentalist semantic theory within an interdisciplinary approach, and if so: how?

In the three preceding chapters several proposals are made for meeting these important objections against John Locke's original empiricist mentalist semantic theory, and as a result of the work reported there we are able to provide the answer to the main research question of this master's thesis. It reads: yes! The objections can indeed be met.

The chapter on intersubjectivity reduces the intersubjectivity objection to a question concerning similarity. Intuitively, two ideas within different subjects are similar if they have the same content, but that content is itself not directly publicly accessible. However, it can be made public in an indirect way with the help of the concept of a 'web-of-ideas'. A web-of-ideas is a collection of ideas that are related to one another, where the relation between ideas reflects their content. In order to account for the similarity of two ideas within different subjects and, hence, different webs-of-ideas, the concept of 'web-similarity' is introduced: two ideas in different webs-of-ideas are web-similar if they occupy the same position within their respective webs. Given this much, the relation between two ideas can be made public by concatenating the – public – signs that stand for the ideas and comparing the concatenations of different subjects.

The chapter on self-applicability discusses two closely related questions that underlie the self-applicability objection against empiricist mentalist semantic theory. Viz.: How are ideas activated? And: What is the nature of ideas? The answers to these questions determine whether ideas can be self-applicable and, hence, whether empiricist mentalist semantic theory is capable of meeting the self-applicability objection. Using the metaphor of an automatic vending machine, the chapter on self-applicability argues that slot-like ideas, in contrast to object-like ideas, are immune to the self-applicability objection. The chapter also discusses what this change in nature entails for their activation process. Rather than being objects that are consciously handled by the subject, ideas are static slots that can become activated in certain types of situations due to the internal structure of the subject, which guides input signals to ideas.

The chapter on abstract ideas discusses two closely related questions that underlie the abstract-ideas objection against empiricist mentalist semantic theory. Viz.: How may ideas represent abstract entities such as ‘triangle’? And: How does the development process of (abstract) ideas proceed? It is shown that the picture-like ideas postulated by Locke cannot represent abstract entities, because they lack the capability of being indeterminate. A more plausible view that does give ideas this capability is developed by combining hints about ideas being pattern-like with the concept of regular expressions. Moreover, the development process of such pattern-like (abstract) ideas is outlined by referring to the various supervised and unsupervised learning techniques within the field of pattern recognition.

Thus, when they are approached from the interdisciplinary field of cognitive artificial intelligence, the three objections may not be as insuperable as they seem. When Locke’s original empiricist mentalist semantic theory is enriched with the modifications proposed in the previous chapters the resulting *updated* empiricist mentalist semantic theory is not only immune to the three objections, but also gives rise to new questions that will be treated in the next section. Before that, we will try and sketch an overview of the updated theory.

The various aspects of the updated theory may best be illustrated on the basis of an imaginary situation. Consider a computer subject that communicates about the presence of a house and tree to a human subject (see Figure 9).

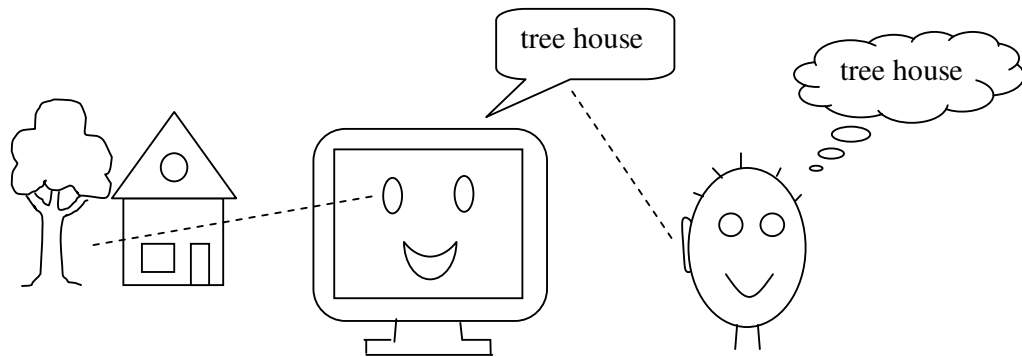


Figure 9

In this situation a tree and a house stimulate the senses of a computer subject. The internal structure of the computer subject guides the signals caused by the tree and the house to the respective slot-like ideas TREE and HOUSE, which are activated. To a subject, the activation of a slot-like idea counts as evidence that some pattern is present in its experience. The internal structure of the computer subject either is the product of a pattern-recognition mechanism that has been programmed into it, or it has been put there by directly programming the desired structure into the computer subject. The content of an idea is determined by its activation pattern and this activation pattern, in turn, is determined by the internal structure. It is possible for ideas such as the ideas for which the signs ‘democracy’ or ‘triangle’ stand to have abstract content, because the activation pattern can be indeterminate. The simultaneous activation of the ideas TREE and HOUSE indicates that there is a relation between them in this specific situation. The activation patterns of all ideas of a subject constitute a web-of-ideas, because they cause ideas to have relations between each other. The content of ideas themselves cannot be made public, such in contrast to the relations between them, which depend on their content. The fact that the ideas TREE and HOUSE are related in the above situation can be made public by the computer subject by uttering the concatenation of the signs for the two related ideas: “tree house”. Thus, the computer subject’s utterance of the concatenation publicizes the part of its web-of-ideas where the idea TREE and the idea HOUSE are related. Since this uttered concatenation is public, it is possible for the other (human) subject to hear it. The concatenation triggers the senses of the human subject and, again, its internal structure conducts the signals to the respective slot-like ideas TREE and HOUSE, which are activated. This simultaneous activation of the ideas TREE and HOUSE suggests that

these two ideas are related in the specific situation and this is what the computer subject has communicated to the human subject.

5.2 Further research

Despite the fact that this master's thesis set out to answer questions, it also raises quite a few questions. This might be seen as evidence that the answers offered are incorrect, but it may also signal that the answers are incomplete. This incompleteness might have several causes. First of all, incompleteness is an inevitable property of answers of science and philosophy. Second, the questions may be beyond the scope of this master's thesis. And third, lack of time may prevent their receiving any attention. In this section some of these questions will be addressed as pointers to further research.

In the previous section it was noted that the content of an idea depends on its activation pattern. (This issue was also addressed in the chapter on self-applicability.) Consider an automatic vending machine with coin-slots that resemble the slot- and pattern-like ideas of empiricist mentalist semantic theory. It has in its internals a coin-slot for 15 euro. Although coins of 15 euro do not exist, the specific coin-slot can be activated by inserting seven 2 euro coins plus one 1 euro coin into the automatic vending-machine. But the 15 euro coin-slot can also be activated by inserting fifteen 1 euro coins or three 5 euro coins (which do not exist) into the automatic vending-machine. So, the automatic vending machine is structured in such a way that the 15 euro coin-slot is activated by every possible combination of coins with a worth of 15 euro. Now, consider the question why this coin-slot is said to be that of 15 euro. The coin-slot is only said to be a 15 euro coin-slot because of the input strings that activate it. Would it be activated by all possible combinations with a worth of 22 euro, then its content would be different and it would be called a 22 euro coin-slot. In an analogous way, a slot/pattern-like idea is said to be the idea TREE, because it is quite reliably activated by trees. Would the idea instead be activated by cars, then its content would be different too. However, specific research on the question whether the content of ideas can be identified with their activation pattern is required for making this claim of empiricist mentalist semantic theory more plausible.

Another question that is raised by the updated theory concerns the claim that the content of ideas is reflected by the web-of-ideas. This question, together with a related question about the exact nature of the web-of-ideas, is also mentioned in the chapter

on intersubjectivity. The proposals in that chapter rely heavily upon the interdependence of the web-of-ideas and the content of ideas. Further research must decide whether this interdependence is possible. Further research may have its starting point in the following. A web-of-ideas can be described as a collection of private ideas within a subject that are related to one another by virtue of their activation pattern. How can the activation patterns of two ideas relate these ideas to one another? Consider the two ideas TREE and TRUNK and their activation pattern. Because a tree always has a trunk, the activation pattern of the idea TRUNK is related to the activation pattern of the idea TREE. The activation pattern caused by a tree always contains the activation pattern caused by a trunk and this can be seen as a relation between the idea TREE and the idea TRUNK. In a similar way the relation between the ideas BLACK and WHITE might be explained. The activation pattern caused by black never contains the activation pattern caused by white and vice versa. This might explain how the ideas BLACK and WHITE are related. Note that the relations between ideas within webs-of-ideas are not the same as their activation relations, although the former depend on the latter. Whether these content-based relations between ideas are powerful enough to explain the wide variety of communication is not clear, but it has at least some capacities to explain relations between them.

Other questions that were already touched upon in previous chapters and that require further research are the following. In the chapter on intersubjectivity parallels were drawn between updated empiricist mentalist semantic theory and Wittgenstein's non-mentalist meaning-is-use semantic theory. It was suggested there that meaning-is-use follows quite naturally from the web-of-ideas approach and that the traditional gap between public and private meaning is bridgeable. More detailed research is required to substantiate these claims. In the chapter on abstract ideas supervised and unsupervised learning techniques from pattern recognition in computer science were mentioned as examples of development processes that are capable of creating pattern recognizers through experience or education. Further research might focus on suitable learning techniques in developing slot- and pattern-like ideas compatible with updated empiricist mentalist semantic theory. The chapter on intersubjectivity described a process capable of synchronizing or converging the webs-of-ideas of a group of subjects. The process explained how web-similarity, and with it intersubjectivity, can arise out of almost nothing. A first step in upgrading these claims from mere

speculation to a real solution might consist of developing a simulation of this process on a computer. This simulation could expose a group of simplified subject with random webs-of-ideas to the conditions sketched in the synchronization process and investigate whether their webs-of-ideas do converge.

Finally we point to further research that was not mentioned earlier in this master's thesis because it falls outside its scope, though it is related and might be worth pursuing. Interesting further research might focus on the debate between empiricists and rationalists. Traditionally, empiricists claim that there are *no* innate principles and rationalists claim that there are *only* innate principles. Updated empiricist mentalist semantic theory allows for the possibility of an innate (pattern-recognizing) mechanism that is responsible for the development process of ideas and may thus strike the golden mean in this historical debate.

Two other topics worth investigating are the nature of signs within empiricist mentalist semantic theory and the parallels drawn between the stand-for relation between signs and ideas and the activation process of ideas. The stand-for relation between signs and ideas elucidates the nature of signs. This stand-for relation is conventional, so that it does not strictly matter which sign is related to which idea, although switching relations might take away the possibility of communication. The stand-for relation, therefore, is an embedded causal relation, because if it would solely depend on the causal laws of nature it could not be conventional, just as the relation between smoke and fire, and water and ice. From the chapter on self-applicability we know that these embedded causal relations are conducted by some structure in reality. This precondition that requires that the stand-for relation be accompanied by a structure between the sign and the idea forces signs to be input patterns to the senses of subjects, for note that if signs would be words, written with ink on some piece of paper, or metal road signs along the road, then there could be no structure in reality between the sign and its corresponding idea. Since the structure of the stand-for relation must still be there when it is not being activated, it is implausible to claim that this structure actually bridges the distance between the piece of paper or the metal road sign and the idea within the subject. (Observe that if the subject is hundreds of miles away from the metal road sign, the structure of the stand-for relation should stretch over hundreds of miles as well.) The problem does not arise with signs that are input patterns, because there the structure of the stand-for relation lies between the

senses of some subject and its ideas. If both the stand-for relation and the activation process of ideas require a structure from the senses to a subject's ideas, then what is the difference between the two (if any)? It may be the case that these relations are the same at a more fundamental level. If that is true, then both the input pattern of the word 'tree' and the input pattern of a real tree travel through the internal structure of the subject to the same idea TREE.

The last topic that did not receive the attention it deserved in this master's thesis concerns the implications of updated empiricist mentalist semantic theory for language. This master's thesis did not consider any interesting part of speech or any non-trivial grammatical construction and restricted itself to very primitive "sentences" consisting of concatenations of only two signs. It did, however, show the necessity of these overly simple concatenations for obtaining intersubjectivity. It might be worth to further investigate the implications, because, although phrases like 'tree house' do not exactly display the fine nuances and flexibility of natural human language in the expression information, they do contain *some* information in a rather simplistic way.

References

- Aaron, Richard I. *John Locke*. Oxford: Clarendon Press, 1955.
- Ayers, Michael. *Locke, Volume I: Epistemology*. London: Routledge, 1991.
- Bennett, Jonathan. *Locke, Berkeley, Hume: central themes*. Oxford: Clarendon Press, 1971.
- Blackburn, Simon. *Spreading the word: groundings in the philosophy of language*. Oxford: Clarendon Press, 1984.
- Bechtel, William, Adele Abrahamsen. *Connectionism and the Mind: An Introduction to Parallel Processing in Networks*. Oxford and Cambridge: Basil Blackwell, 1991.
- Candlish, Stewart. *Private Language*. The Stanford Encyclopedia of Philosophy (Spring 2004 Edition), Edward N. Zalta (ed.).
<http://plato.stanford.edu/archives/spr2004/entries/private-language/>.
- Chang, Chen Chung, H. J. Keisler. *Model Theory*. Elsevier Science Publ., 1990.
- Cook, John W. "Wittgenstein on Privacy." *The Philosophical Review* 74.3 (July 1965): 281-314.
- Cooper, David E. *Philosophy and the nature of language*. London: Longman, 1975.
- Crossley, Nick. *Intersubjectivity: the fabric of social becoming*. London: Sage, 1996.
- Currie, Gregory. *Frege: An introduction to his philosophy*. Harvester Press, 1982.
- Duda, Richard O., Peter E. Hart and David G. Stork. *Pattern Classification*. Wiley-Interscience, 2001.
- Dunn, John, James O. Urmson, Alfred J. Ayer. *The British Empiricists: Locke, Berkeley, Hume*. New York: Oxford University Press, 1992.
- Fodor, Jerry. *The Language of Thought*. New York: Harvard University Press, 1975.
- Friedl, Jeffrey. *Mastering Regular Expressions: Powerful Techniques for Perl and Other Tools*. Tokyo: O'Reilly, 1997.
- Hacking, Ian. *Why does language matter to philosophy?* Cambridge University Press, 1975.
- Hendriks, Petra, Niels Taatgen and Tjeerd Andringa. *Breinmakers & Breinbrekers: Inleiding cognitiewetenschap*. Netherlands: Addison Wesley Longman, 1997.
- Jolley, Nicholas. *Locke: his philosophical thought*. Oxford: Oxford University Press, 1999.

- Locke, John. (1690) *An Essay Concerning Human Understanding*. Ed. Nidditch, Peter. London: Oxford University Press, 1975.
- Lowe, Jonathan. *Locke: On Human Understanding*. Routledge Philosophy GuideBooks, 1995.
- Margolis, Eric, and Stephen Laurence, eds. *Concepts: Core Readings*. Cambridge: The MIT Press, 1999.
- Martinich, Aloysius P, ed. *The philosophy of language*. New York: Oxford University Press, 1996.
- O'Connor, D.J. *John Locke*. New York: Dover Publications, 1967.
- Picardi, Eva. "Frege's anti-psychologism." *Gottlob Frege*. Critical assessments of leading philosophers. Ed. Beaney, Michael, and Erich H. Reck. London: Routledge, 2005.
- Rabb, J. Douglas. *John Locke on Reflection: A Phenomenology Lost*. Washington, D.C.: The Center for Advanced Research in Phenomenology & University Press of America, 1985.
- Roeffaers, Hugo. *Taal en woordkunst: een filosofische verkenning*. Leuven: Garant, 1997.
- Stokhof, Martin. *Taal en Betekenis: Een inleiding in de taalfilosofie*. Amsterdam: Boom, 2003.
- Wittgenstein, Ludwig. (1953) *Philosophical Investigations*. Trans. G.E.M. Anscombe. 3rd ed. Oxford: Blackwell, 2001.
- Wittgenstein, L. *The Blue and Brown Books*. Oxford: Blackwell, 1958.
- Woudenberg, René van. *Filosofie van taal en tekst*. Budel: Damon, 2002.
- Yolton, John W., *Locke: An Introduction*. Oxford: Blackwell, 1985.