

The Emergence of Meaning

Peter Gärdenfors
Cognitive Science
Department of Philosophy
Lund University, Sweden

1. The problem of social meaning

When we communicate by words we assume that there is something, a language, that forms a sufficient basis for the success of the communication. It is not enough for communication that the speakers use the same syntax and vocabulary – they must *mean* the same things as well, i.e., they must also have a common semantics.

It is commonplace that language is conventional in the sense that the connections between many of the linguistic signs and their meanings are arbitrary. Consequently, the semantics of a language has to be *learned* by individual speakers – the innate aspects of language acquisition do not include the semantic mapping between words and their meanings. However, even if it is assumed that all speakers use the same vocabulary and grammar, it remains to be explained how they can mean the same thing, i.e., that there is some well-behaved relation between individual uses of language and the social meaning. The task of this paper is to provide an architecture for such a relation. My discussion will focus on the following two questions:

Question 1: What exactly is the relation between individual speakers and their communal language?

Question 2: Where are these things called the meanings to be found?

The main thesis of this article is that the social meaning *emerges* from individual meanings. As an answer to question 1, I shall argue that, apart from individual meanings, the factor that determines social meaning is the structure of *linguistic power* in the society. The emerging social meaning will be exhibited with the aid of some elementary tools from model theory.

I shall also argue that the emergent social meaning has certain causal powers. In particular it functions as a *regulative ideal* for individual assignments of meaning. Like money, language is a *social good*. Furthermore, social meaning is the point of departure for an explanation of the *conventional* nature of language.

2. Two analogies for the emergence of social meaning

Before I start elaborating my arguments for meaning as an emergent phenomenon in a society of language users, I want to illustrate the general idea of an emergent property by two examples.

The concept of an emergent property of a system is a fairly old idea within cybernetics which is perhaps best illustrated by Norbert Wiener's (1961) famous example of the 'virtual governor': Consider a system that consists of a network of AC generators. Each generator has built into it a regulator that controls its speed so that it deviates very little from 60 Hz at any time. However, a generator in isolation does not give a very steady 60 Hz output. In remarkable contrast, when a large number of such generators are interconnected, they behave much more stably. This 'mutual entrainment' of the generators is an example of *self-organization*. Out of the mutual entrainment emerges what Wiener calls a 'virtual governor' which is an *equilibrium* property of the entire system that is viewed as having *causal effects* on the individual generators in the system. If a new generator is added to the system, the effects are best explained by saying that the virtual governor causes it to 'get into step' by pumping energy into it, if it lags in phase, or by absorbing energy, if it runs too fast.¹

The second example concerns the emergence of a *price vector* in an equilibrium exchange market. When describing a market exchange economy some idealized assumptions about the agents are made. The assumptions are formulated in terms of *preferences* and the initial endowments of *goods*. For example, preferences are assumed to be continuous and decreasing so that more of a good is always preferable to less.

The prices of goods are not mentioned in these assumptions, nor is the concept of price presumed. Nevertheless, it is an important result about an exchange economy in equilibrium that the individuals in the market will behave *as if* there exists a set of fixed prices, i.e., a price vector, for the goods, even though there is no such thing as money available in the market. As a matter of fact, without something close to such an equilibrium, money cannot exist. The price vector is an emergent phenomenon that acts as a virtual governor for the exchanges on the market. The existence of a (relatively stable) price vector is a necessary prerequisite for monetary *conventions* to develop.² When such conventions evolve, money becomes a *social good*.³

An idealizing requirement which is necessary to prove the *uniqueness* of the emerging price vector is that there are *infinitely* many agents in the economy. If there were only finitely many, each of the agents could change the price vector by altering his preferences. This assumption is, of course, unrealistic, but it can be approximated to an arbitrary degree by assuming a

¹For an analysis of this example and a discussion of 'consciousness' as a virtual governor of the brain, cf. Dewan (1976).

²For a fascinating account of how a monetary system develops and breaks down in a prisoners of war camp, cf. Radford (1945). He also presents an analysis of the conditions under which the necessary conventions may function.

³For an elaboration of this example, see Gärdenfors (1990).

sufficiently large number of agents. However, the assumption considerably simplifies the mathematical techniques of the proof.

These two examples are put forward as mere *analogies* to my thesis that the meaning of linguistic expressions in a society is something that emerges out of the individuals' conceptual mappings and the underlying structure of linguistic power. In order to turn this analogy into something more substantial, the main task is to show how the emergent social meaning can be determined. I shall argue that the social meaning is crucially dependent on some features of linguistic power. However, before I can present my analysis, we must put the problem of social meaning in a broader semantic context.

3. "When *I* use a word," Humpty Dumpty said, in rather a scornful tone, "it means just what I choose it to mean – neither more nor less."⁴

A semantic theory can only answer questions like 1 and 2 above in combination with answers to the following:

Question 3: What determines the meaning of the expressions of a language?

Question 4: How can individual users *grasp* the meanings of expressions?

Regarding the answers to the latter questions, I believe one can roughly distinguish two prevailing traditions in semantics, one *realistic* (or *externalistic*) and one *conceptualistic*.

According to the realistic approach to semantics the meaning of an expression is something out there *in the world*. In technical terms, a semantics for a language is seen as a mapping from the grammatical structures to things in the world (or in several possible worlds).⁵ Often meanings are defined in terms of *truth conditions*. A consequence of this approach is that the meaning of an expression is independent of how individual users understand it.

For this approach the answers to questions 2 and 3 are immediate: It is the semantic mapping that determines meanings. However, questions 1 and 4 comprise problems for semantic realism since the semantic mapping does not tell us anything about how individual users 'grasp' the public meanings.

The second paradigm of semantics I want to focus on is conceptualistic or cognitivist. The central tenet of this approach is that meanings of expressions are *mental entities*. A semantics is seen as a mapping from the linguistic expressions to cognitive structures. The external world enters on

⁴This quotation, as well as the following two section headings are taken from Carroll (1871).

⁵According to so called causal theories of reference, the fundamental parts of the mapping between the language and the world consist of causal relations which "hook" some of the expressions of a language to their external references.

the scene only when the relation between it and the cognitive structure is considered. According to this kind of semantic theory the relation between meanings and the external world is secondary, and only determined after the cognitive structures have been settled. As a consequence, *meaning becomes independent of truth*.

For the conceptualistic approach, question 4 is not a problem: A person grasps the meaning of an expression by relating it to a cognitive structure. However, we will now have trouble with questions 1 and 3. If everyone determines meanings in his own head, how can we talk about *the* meaning of an expression in a society of language users? A similar question is: How can we say that someone is *wrong* about the meaning?

The answer to question 2 is given by the prime slogan for cognitive semantics: *Meanings are in the head*. More precisely, a semantics for a language is seen as a mapping from the expressions of the language to some cognitive or mental entities. Langacker (1986a, p. 3) formulates it crisply, "Meaning is equated with conceptualization." On this approach, the way an individual *grasps* the meaning of an expression cannot be separated from the process of *assigning* meaning to the expression. In other words, the cognitive approach refuses to divorce semantics from psychology. The semantic mapping is still *arbitrary* in the sense that any speaker can assign any particular linguistic expression to represent any cognitive structure (thus Humpty Dumpty). In practice, however, the choices are governed by a system of social *conventions*.

Since the cognitive structures in our heads are connected to our perceptual mechanisms, directly or indirectly, it follows that *meanings are*, at least partly, *perceptually grounded*. This, again, is in contrast to traditional realistic versions of semantics which claim that since meaning is a mapping between the language and the external world (or several worlds), meaning does not depend on perceptual structures.⁶

Jackendoff (1988, p. 81) argues that the consequences of adopting a cognitivist semantics flow in two directions:

⁶To be sure, in a causal theory of reference (like Kripke's), a causal relation determining the reference of some expression can be established via perception. However, such a causal link is not dependent on the cognitive structure of the perception.

"On the one hand, research on the nature of human mental representation, independent of issues of meaning, can be used to constrain or enrich semantic theory; on the other hand, results in semantic theory can be taken to bear directly on questions of human conceptualization. Thus ... semantic theory no longer is just an aspect of the study of language (or of logic). Rather, it becomes an element of a wider theory of psychology, fully integrated into the study of mind".

Cognitive semantics should be separated from Fodor's (1981) "language of thought" hypothesis. There are similarities, though: Fodor also uses mental entities to represent linguistic information. This is his 'language of thought' which is sometimes also called 'Mentalese'. According to Fodor, this is what speakers use when they compute inferences (according to some internal set of rules) and when they formulate linguistic responses (translated back from Mentalese to some appropriate natural language). However, the mental entities constituting Mentalese form a *language* with syntactic structures governed by some recursive set of rules. And when it comes to the *semantics* of Mentalese, Fodor is still a realist and relies on references in the external world and truth conditions.

In contrast, the mental structures applied in cognitive semantics *are* the meanings of the linguistic idioms; there is no further step of translating conceptual structure to something outside the mind. Furthermore, instead of being a symbolic system having syntactic structure like Mentalese, the conceptual schemes that are used to represent meanings are often based on *geometric* or *spatial* constructions. Prime examples of such constructions are Lakoff's (1987) 'image schemas' and Langacker's (1986a,b) semantic diagrams. Related versions of cognitive semantics can be found in the writings of Jackendoff (1983, 1988), Johnson-Laird (1983), Talmy (1989), Gärdenfors (to appear) and many others.

4. "The question is," said Alice, "whether you *can* make words mean so many different things."

But, if everybody can mandate his own cognitive meaning, how can we then talk about *the* meaning of an expression? And how can somebody be *wrong* about the meaning? If cognitive semantics with its emphasis on individual conceptual structures is correct, why do we not have Babel?

For a realist semantics these questions pose no serious problems since such a semantics assumes that meanings exist in the external world and not in our heads. People are wrong about the meaning of an expression whenever they have not grasped the correct 'external' meaning of it.

Realists tend to eschew questions like these and 1 and 4 above by driving a wedge between semantics proper and psychological and sociological analyses of the use of language. Thus Lewis (1970, p. 19):

"I distinguish two topics: first, the description of possible languages or grammars as abstract semantic systems whereby symbols are associated with aspects of the world; and second, the description of the psychological and sociological facts whereby a particular one of these abstract systems is the one used by a person or population. Only confusion comes of mixing these two topics."

From a realist perspective it seems incoherent to say that the meaning of a word for one individual is different from the meaning for another. This can only be the case when someone has not understood (or grasped) the correct meaning of a word.

Much of Tarski's model theory and Davidson's truth conditions program are motivated by this idea. It should be noted, however, that model theory *per se* does not presume realism but is compatible with a conceptualistic semantics since it is possible to construct a model that conforms to cognitive constraints. In a sense, one of the goals of the project of cognitive semantics can be described as an endeavor to produce a *mental* model theory (cf. e.g. Johnson-Laird (1983)).

Another attack against cognitive semantics has been launched by Putnam (1975, 1988). His argument that meanings can't be in the head starts from the following assumptions about meaning and mental representations, all of which are accepted by the cognitive semanticists (Putnam 1988, p. 19):

1. Every word the speaker uses is associated in her mind with a certain mental representation.
2. Two words are synonymous (have the same meaning) just in case they are associated with the *same* mental representation by the speakers who use those words.
3. The mental representation is what the word refers to if anything.

Putnam claims that these three conditions cannot be simultaneously satisfied. The reason is that we "cannot individuate concepts and beliefs without reference to the environment" (1988, p. 73). A central part of his argument can be illustrated by the following example (Putnam (1975), p. 226-227):

"Suppose you are like me and cannot tell an elm from a beech tree. We still say that the extension of 'elm' in my idiolect is the same as the extension of 'elm' in anyone else's, viz., the set of all *elm* trees, and that the set of all beech trees is the extension of 'beech' in both of our idiolects. Thus 'elm' in my idiolect has a different extension from 'beech' in your idiolect (as it should). Is it really credible that this difference in extension is brought about by some difference in our *concepts*? My *concept* of an elm tree is exactly the same as my concept of a beech tree (I blush to confess). (This shows that the identification of meaning 'in the sense of intension' with *concept* cannot be correct, by the way). ... Cut the pie any way you like, meanings just ain't in the *head*!"

The upshot is, realists would claim, that meanings must refer to something non-cognitive, and thus Humpty Dumpty cannot really mean at his whims.

5. "The question is," said Humpty Dumpty, "which is to be master – that's all."

All semantic theories agree that meaning is conventional to some extent since a unanimous society can decide to use any particular word in any way they want. Even if different theories give different answers to question 2 above, none of them, to my mind, answers question 1 satisfactorily. Once we focus on the relation between individual and social meanings, the following becomes an inexorable problem for a semantic theory: *Who* decides on what is the correct meaning of an expression in a society? This question must be answered independently of whether one is a cognitivist or a realist.

Normally, a single speaker does not have the power to set the meaning of a word or expression. A marginal case is the use of *stipulative definitions* in science or in law. An author of an algebraic text can for example decide to use the word 'filter' to denote a collection D of subsets of a set U such that (i) $U \in D$; (ii) $\emptyset \notin D$; (iii) if $X \in D$ and $Y \in D$, then $X \cap Y \in D$, and (iv) if $X \in D$ and $X \subseteq Y$, then $Y \in D$ (this definition will come in handy later on).

Instead of answering the question of who is the master of meaning in a sociolinguistic fashion, I shall formulate some general principles concerning linguistic power and show how such principles will *constrain* the command of meaning. The point of this is to show that even very general requirements for the social structure of meaning will have effects on the possible patterns of linguistic power.

In order to procure some precision in the results, the conditions will be fitted with a formal dress. Let L be a language and M be a set of 'meanings' Let L_0 be the set of 'atoms' of L . The nature of the elements in L_0 is left open for the time being; they can be words, phrases, or sentences. Most of the time, though, I will write as if they are sentences. The elements of L will be denoted a, b, c, \dots .

Primarily, the 'meanings' in M are intended to be *propositions*, i.e., meanings of sentences, and my choice of terminology will, to some extent, reflect this interpretation. However, I want to emphasize that there is nothing to prevent that they are interpreted as *concepts*, i.e., meanings of single words. The only assumption concerning the structure of M that will be made at this stage is that there is a *logic* defined on M ; $m_i \vdash m_j$ denotes that m_i logically entails m_j . (If meanings are concepts, $m_i \vdash m_j$ denotes that the meaning of m_j is included in the meaning of m_i). The reason for keeping the nature of meanings unspecified is that I want to develop conditions on linguistic power which are as general as possible and thus neutral with respect to the issue of realism vs. conceptualism.

Let U be a society of *users* of L with individual members i, j, \dots . An *individual semantics* is a mapping m_i from L to M , where $i \in U$. A *social semantics* is a mapping m_S from L to M . A *semantic situation* S is a set of individual mappings, one for each $i \in U$. In other words, the actual semantic situation is a function of the individual mappings m_i , so that if any of these mappings changes, the situation changes too.

Using this technical framework, I shall now present two sets of general conditions on the relation between the individual semantic mappings m_i and the social mapping m_S . The first set of requirements, the A-conditions, is associated with a particular view of meaning. According to this view, what matters for meanings are their *logical relations* only. In other words, the meaning of an expression is uniquely delineated when its logical relations to other meanings are decided.⁷

The first condition in this set requires that each speaker masters his own semantics in the sense that, for any meaning, he can choose an arbitrary atomic expression to denote that meaning.

(A1): (*Arbitrariness of linguistic sign*) Any mapping from L_0 to M is a possible individual semantics, i.e., for any $a \in L_0$, $m_i(a)$ can be chosen to be an arbitrary element of M .

The intention behind the second condition is that if all members of U in a given semantic situation S agree upon the meaning of an expression, then that meaning is indeed the social meaning in S of the expression. However, given the assumption that logical relations are what matters and nothing else, only the following weaker requirement is introduced:

(A2): (*Logical Unanimity*) If for every $i \in U$ in a semantic situation S , $m_i(a) \vdash m_i(b)$, then $m_S(a) \vdash m_S(b)$.

The next condition concerns the properties of the logical relation \vdash . The only properties that are needed for my purpose are:

(A3): (a) (*Transitivity*) The relation \vdash is transitive; (b) (*Non-triviality*): For some a and b it does *not* hold that $m_S(a) \vdash m_S(b)$.

The final condition is the most difficult to defend. It formulates a strong version of the assumed theory of meaning. The condition requires that as soon as nobody changes his opinion concerning the logical relation between two expressions a and b , the logical relation between a and b in the social meaning will remain the same. In other words, the logical relation between $m_S(a)$ and $m_S(b)$ is *independent* of the meanings of all other expressions.

(A4): (*Logicity of meaning*) If, for all individual mappings m_i , the logical relations between $m_i(a)$ and $m_i(b)$ are the same in situations S and S' , then the logical relations between $m_S(a)$ and $m_S(b)$ on the one hand and $m_{S'}(a)$ and $m_{S'}(b)$ on the other are the same.

I do not want to make any efforts to furnish deeper motivation for these conditions (simply because I do not think such a defense exists). The conditions should rather be seen as *examples* of what can be said about linguistic power in general terms. Still it is interesting to see that conditions (A1) - (A4) have some non-trivial consequences concerning who masters meaning.

⁷Frege comes close to this view and Dummett maybe even closer.

Let us say that a set D of individuals, $D \subseteq U$, is *decisive for L* if, for any situation S , it holds that $m_S(a) \vdash m_S(b)$, whenever $m_i(a) \vdash m_i(b)$ for all $i \in D$. In other words, if the speakers in D agree upon the logical relations of the meanings of a and b , then these relations hold for the social meanings too. If the group D is decisive for L , then D can be said to master the social meaning of L . There may, of course, be more than one decisive set in a society.

The ground is now prepared for the following result where we look at the set of all decisive sets.

Theorem 1: If conditions (A1) - (A4) are satisfied, then the set D of decisive sets of individuals forms a *filter*.

The proof of the theorem will be found in the appendix.⁸ It is adapted from a proof of Arrow's impossibility theorem due to Hansson (1976). As a matter of fact, I can now reveal that the conditions (A1) - (A4) have been formulated to parallel the conditions of Arrow's theorem. (A4) corresponds to his 'independence of irrelevant alternatives'. The role of the preference relations in his conditions is played by the relation \vdash in the present setting.

There is one important difference though. Arrow assumes that the preference relation is *connected*. The corresponding assumption for \vdash , i.e., that either $x \vdash y$ or $y \vdash x$ for any x and y , is of course absolutely preposterous. However, had it been added to the other conditions, we could have shown that the set of decisive sets forms an *ultrafilter*.⁹ This result would have given the full power of Arrow's theorem, since an ultrafilter on a finite set always contains a decisive singleton, i.e., a *dictator*.

I am not impressed by the theorem as a result about semantics, since condition (A4) is much too *ad hoc* and introduced mainly to fit with the structure of Arrow's impossibility result. Notwithstanding, the theorem shows that even some very general conditions on the interaction between individual meanings and social meaning can *constrain* the possible ways social meaning is determined.

The conditions that have been utilized for Theorem 1 make hardly any assumptions about the *structure* of the language L or the set of meanings M . For the second set of conditions on semantic power, the B-conditions, we will impose some structure on these sets by assuming that both the language L and the set M of meaning are closed under *Boolean operators*. More precisely, L is supposed to be closed under the standard sentential connectives \neg , $\&$, and \vee , with \rightarrow and \leftrightarrow defined as usual.¹⁰ In addition M is supposed to be a *Boolean algebra*. By definition, such an algebra is closed under complementation $*$, meet \cap and join \cup (corresponding to \neg , $\&$, and \vee

⁸For a definition of 'filter' see above.

⁹ Cf. Hansson's (1976) proof.

¹⁰Again, if the atomic expressions of L are thought of as predicates rather than sentences, the connectives can be interpreted as operators on predicates.

respectively). The unit and zero elements of the algebra (corresponding to a tautology and a contradiction respectively) will be denoted \top and \perp .

Assuming this structure, it is then possible to formulate the principle of *compositionality* of meaning in a precise way. This principle, which is due to Frege and is one of the cornerstones of Montague-type semantics, states that the meaning of a composite expression is *determined* by the meanings of its constituents. Here I only need compositionality with respect to Boolean operators. Algebraically, it corresponds to the following condition:¹¹

(B1) (*Boolean compositionality*) The mappings m_i and m_S from L to M are *homomorphisms*, i.e., for all a and b it holds that $m_i(\neg a) = m_i(a)^*$, $m_i(a \ \& \ b) = m_i(a) \cap m_i(b)$, and $m_i(a \vee b) = m_i(a) \cup m_i(b)$, and similarly for m_S .

A consequence of this condition is that the connectives are given their classical truth-functional meaning.

The second condition concerns the *analyticity* of expressions. The idea is that if everyone in the society assigns a contingent meaning to an expression a , i.e., a is considered by everyone to be neither a tautology nor a contradiction, then the social meaning of a is contingent too.

(B2) (*Contingency preservation*) If for all $i \in U$ in a semantic situation S , $m_i(a) \neq \top$, then $m_S(a) \neq \top$.

The dual condition that if $m_i(a) \neq \perp$ for all m_i in S then $m_S(a) \neq \perp$ follows easily from (B2) together with (B1).

The third and final condition I shall need requires that there be some means for designating or naming social meanings in the language. As a motivation for the assumption, it can be noted that it is common to speak of 'the meaning' of an expression, where the intention is precisely to refer to the social meaning of it.¹² For many kinds of formal semantic theories, the existence of such designators comes quite naturally. For example, consider a possible worlds semantics where the elements of L are sentences and the elements of M representing the meanings of the sentences are *sets of possible worlds*, i.e., 'propositions'. In such a semantics a designator of a proposition would simply be an extensional description of the set of possible worlds that is identified with the proposition.

(B3) (*Designators for social meaning*) For all a in L and all semantic situations S , there exists an expression a^S in L such that $m_i(a^S) = m_S(a^S) = m_S(a)$ for all $i \in U$.

¹¹Cf. van Benthem (1986, p. 200) for an analogous definition. Interestingly enough, he concludes that (in the individual setting) "compositionality provides no significant constraint upon semantic theory." Theorem 2 of this paper is an example of the converse (in the social setting).

¹²Also 'hedges' (cf. Lakoff 1987) like "strictly speaking" and "technically" have a similar function.

Note that individual i need not be able to *identify* the reference of $m_i(a^S)$; she is only assumed to know that there is a designator of the social meaning. Furthermore, in the presence of compositionality it is sufficient to assume that the atomic expressions of L_0 have such designators. Designators for complex expressions can then be constructed from these primitive ones.

Most of the time, I write as if the atoms of L are sentences. As a counterbalance, I will now present an example where the constituents of L are *predicates*. The point of the example is to illustrate how all conditions (B1) - (B3) can be fulfilled simultaneously. So assume that all atomic a in L denote *properties*. The Boolean operators are now interpreted as operators on predicates so that if a denotes "mother" and b "father", then $a \vee b$ denotes "parent" etc.

The meanings in M are supposed to be generated from a *finite* set $O = \{o_1, o_2, \dots, o_m\}$ of objects so that $m_i(a)$ and $m_S(a)$ are always subsets of O . Formally, we identify M with the power set of O . The Boolean operators on meanings then become ordinary set-theoretical operators on O , which means that for all semantic mappings $m(a)^* = O - m(a)$ etc. The meanings of the predicates are thus of a familiar *extensional* kind.

In order to manage (B3), it is assumed that for each o_k in O , there is an *atom* a_k in L such that $m_i(a_k) = o_k$, for all individuals i in U . Thus a_k is a designator for a social meaning denoting the property of being identical with o_k . Furthermore, in accordance with compositionality, we let $a_k \vee a_j$ be the social designator for the set $\{o_k, o_j\}$; and so on for all subsets of O . Since O is finite we can in this way construct social designators for all possible 'meanings', i.e., subsets of O . Finally, suppose that we are in a given situation S with individual mappings m_1, \dots, m_n . If 1 is the first individual in a preset ordering of U , we define the social mapping by putting $m_S(a) = m_1(a)$, for all a in L . Thus the social semantics in S in this example is identical to 1's semantics.

It is now trivial to show that for this particular model of the semantic mappings, all conditions (B1) - (B3) are satisfied. (B2) is trivially satisfied by the special choice of social semantics. Apart from being an illustration to how (B1) - (B3) work, this example also provides a verification of the *consistency* of this set of conditions.

Even if each of the conditions (B1) - (B3) seems reasonable, they jointly have far-reaching consequences. In the example just given the social semantics was identified with the semantics of one of the individuals. I will now show that it is not a coincidence. To express this feature more succinctly, let us say that an individual i is a *semantic arbiter* in the situation S iff $m_S(a) = m_i(a)$ for all a in L .

Theorem 2: Assume that U is finite. If the semantic mappings in a situation S satisfies (B1) - (B3), then there exists a semantic arbiter in S .

The proof will be found in the appendix. The idea of the proof is quite simple: If no one is an arbiter, then there is for every speaker i some

expression a_i such that i 's meaning is different from the social meaning of a_i . Since the social meaning of a_i has a designator in L it is then always possible to construct an expression in L which, by compositionality, is contingent for all speakers, but which will turn out to be a tautology in the social meaning, thus violating the condition of contingency preservation.

The earlier definition of a decisive set of individuals was dependent on the choice of the condition (A4). A superior definition that is appropriate in the present context is to say that a set D of speakers, $D \subseteq U$, is *decisive for L in S* if it holds that $m_S(a) = m_i(a)$, whenever $m_i(a) = m_j(a)$ for all $i, j \in D$. Since the set D of all sets that include i form an *ultrafilter*, another way of formulating Theorem 2 is to say that if U is finite, then the set of decisive sets in a situation S that satisfies (B1) - (B3) forms an *ultrafilter*.

None of the conditions (B1) - (B3) say anything about the relations between meanings in *different* situations. In principle, different individuals could be semantic arbiters in varying situations. What we really want to know is what sets of individuals are decisive in *all* semantic situations. It would be interesting to know whether one could deduce something concerning the relations between the arbiters in different situations by adding some condition on social meaning that relates these situations, for example, like what is stipulated in Arrow's independence condition.

6. Democratic vs. oligarchial power structures

Even if the premises of the two theorems above are debatable, their conclusions have the common feature that they suggest that *filters* and *ultrafilters* are interesting representations of semantic power structures. In this section, the role of such structures in determining the social meaning will be further illustrated.

Take a (small) subset D of U . The set \mathbf{D} of all subsets of U that contain D forms a filter. A filter which can be construed in this way is called the *principal filter generated* by D . In the present context such a filter corresponds to an *oligarchial* power structure over linguistic meaning. If D contains only one individual, we obtain an ultrafilter, which then corresponds to a *dictatorial* mastery of meaning.

The oligarchial or dictatorial power structure would emanate in a society where the social meanings are determined by a group of linguistic *experts* writing dictionaries, encyclopedias, handbooks on the proper use of the language, etc. When a member of such a society is in doubt about the meaning of a locution he would rely on the judgments of these experts.

If U is finite, all filters are principal. This means that in a finite society, all semantic power structures are oligarchial or dictatorial. However, if we, like in exchange economics, make the idealizing assumption that U is infinite, there are other kinds of filters. In this context, an interesting example is the so called *Fréchet filter* \mathbf{F} which is defined as the set of all sets F such that $U-F$ contains only finitely many speakers. Such a set F is called *cofinite*. It is easy to show that the intersection of two cofinite sets is also cofinite.

Intuitively, a cofinite set F contains 'almost all' individuals in U . The power structure determined by a Fréchet filter corresponds to a *democratic* power structure since almost everyone must agree on the meaning of an expression if it is to be determined as the social meaning.¹³

A semantic power structure like the one determined by the Fréchet filter is compatible with the conditions (A1) - (A4) used for Theorem 1 since that theorem holds also for an infinite set of individuals.¹⁴ Such a power structure would appear in a society where linguistic meaning is identified with 'common usage'. In a society with this power structure a dictator or a small group of speakers cannot, by themselves, change the meaning of an expression; for this, the consent of almost all language users is mandatory. This is analogous to prices in a free market – a single agent cannot decide to change the price of a good (cf. the example in Section 2).

I do not claim that all parts of the semantics of a language is governed by the same power structure. A more realistic description is to say that a language is a conglomerate of several *sublanguages*, each with its own conditions of mastery. The semantics of the language of lawyers is determined by criteria

¹³ Note that the set of majorities of U does not constitute a filter since the intersection of two majorities need not be a majority.

¹⁴ I do not know whether it is compatible with the conditions (B1) - (B3).

that are different from those of the language of entomologists; which in turn are different from the canons used for slang expressions. For lawyers' and entomologists' expressions the power structures may very well be oligarchial, while the use of slang is a more democratic business. In support of this, it seems as if the 'hedge' "technically" can be used for expressions that are governed by an oligarchial power structure, but not for those the meaning of which are determined democratically: "Technically, a spider is not an insect" is correct, but "Technically, a hooker is a prostitute" sounds odd.¹⁵

Putnam (1975, pp. 227-229) describes something very much like an oligarchial power structure in his hypothesis about the 'division of linguistic labor'. This hypothesis maintains that every linguistic community "... possesses at least some terms whose associated 'criteria' are known only to a subset of the speakers who acquire the terms, and whose use by the other speakers depends upon a structured cooperation between them and the speakers in the relevant subsets" (p. 228).

He claims that the hypothesis accounts for the failures of the assumptions that the knowing the meaning of a locution is just a matter of being in a certain psychological state and that the meaning of a term determines its extension. Putnam's argument for this is that "[w]henver a term is subject to the division of linguistic labor, the 'average' speaker who acquires it does not acquire anything that fixes its extension. In particular his individual psychological state *certainly* does not fix its meaning; it is only the sociolinguistic state of the collective linguistic body to which the speaker belongs that fixes the extension" (p. 229). The last remark indicates that Putnam thinks of the fixation of social meaning in much the same way as in the present article. However, it seems as if he misses the possibility of *democratic* power structures, which is a different way of determining social meaning.¹⁶

In a more recent book, Putnam (1988) also discusses 'conceptual role' semantics, in particular in relation to *natural-kind terms*. He argues that the meaning of such terms cannot be given in terms of their conceptual roles only, but "once we have identified a word as a natural-kind term, then we determine whether it is synonymous with another natural-kind term primarily on the basis of the extensions of the two words" (p. 50). Here, *extension* is, of course the set of things in the world that the word applies to. So natural-kind terms presume a *realistic* component for their semantics according to Putnam.

But, how do we *know* when something is a natural-kind term? Putnam is aware of the problem:

"Some words which were intended to be natural-kind terms turn out not to refer to natural kinds. "Phlogiston" was intended to be the name of a natural kind, but it turned out that there was no such natural kind. And similarly for "ether" and "caloric". In these cases it

¹⁵Cf. Lakoff (1987), pp. 122-125.

¹⁶ The only remark in this direction is the following: "It would be of interest, in particular, to discover if extremely primitive peoples were sometimes exceptions to this hypothesis (which would indicate that the division of linguistic labor is a product of social evolution), or if even they exhibit it" (p. 229).

does seem that something like conceptual role is the dominant factor in meaning, for obvious reasons; we don't want to say that the words "ether" and "caloric" and "phlogiston" are synonymous just because they have the same (empty) extension. ... Indeed the conceptual role theory comes closest to being true in the case of words with an empty extension." (p. 50)

However, here he seems to rely on some form of *realist essentialism*. If "phlogiston" could turn out not to be a natural-kind term, so can "water" and "gold", unless one assumes that natural kinds exist *independently* of language and cognition. And this is the kind of essentialist assumption Putnam needs to make in order to argue against the conceptual approach. But this is putting the cart before the horse: He assumes that a term is defined in terms of realist notions in order to show that it cannot be given a purely conceptual meaning.

In contrast to Putnam I claim that the meaning of natural-kind terms like 'gold' and 'water' *do* change because of changes in the linguistic power structure. Orwell's "Newspeak" is a fictionary example of this phenomenon. I believe that this kind of meaning change is common in science in connection with scientific revolutions. For example, before the Copernican revolution "the earth" *meant* something that did not move, and before Einstein "mass" was something that was constant of an object.¹⁷

7. An unadorned model theoretic construction of social meaning

It is now time to turn to the task of showing how a semantic power structure molds the emerging social meaning. My tools for the construction will be taken from model theory. Since my general approach to semantics is cognitivistic, I would have used a cognitivistic model theory, if a developed theory of that kind had existed. There are some attempts in the direction of a cognitivistic model theory in the writings of Lakoff and Langacker among others. But since it is not presented in sufficient rigor yet, I will work with traditional Tarskian models. Even though such models are normally associated with a realistic approach to semantics (where the models are interpreted as models of a mind-independent reality), it should be noted that there is nothing in the formalism that prevents a *cognitivist interpretation* of the Tarskian type of models, for example by assuming that the elements in the domain of the model refer to cognitive entities rather than to things in the 'external world'.¹⁸ Accordingly, the model structures to be used here are intended to represent individual *mental* models of meaning. I do not pretend that it is a justifiable representation, but if we strive for stringency, it is the best there is.

So we assume that L is a first order language with predicates P_j and constants b_k . We also *assume* that for each $i \in U$, M_i is an appropriate Tarskian model structure for L . The M_i s can be defined as triples $\langle E_i, \{R_{ij}\}, \{c_{ik}\} \rangle$, where E_i is a set of entities, the 'universe', and where the R_{ij} 's are relations between (or

¹⁷For examples in the social sciences, cf. Foucault's "archeological" investigations of the terminologies in various areas ("madness" for instance).

¹⁸If one wants to account for intensional features of language, Montague-style models can be used for the same purpose, except that the definitions will become trickier.

properties of) these entities, matching the predicates P_j of L , and the c_{ik} s are elements of E_i , assigned to the constants b_k of L . For each individual $i \in U$, the mapping m_i from L to M_i is now defined in the standard way.

Now if D is a filter (or an ultrafilter) of decisive sets of individuals we can construct the *reduced product* (or *ultraproduct*) $\prod M_i/D$ of the individual models. This product is well-defined and always a model structure of the right type for L .¹⁹ The construction of a reduced product, which can be found in any good textbook on model theory, starts from the cartesian product $\prod E_i$ of the universes of the models M_i . An element of this product is a sequence $e = \langle \dots, e_i, \dots \rangle$, where $e_i \in E_i$. On elements e and f of the Cartesian product a relation \approx_D is defined so that $e \approx_D f$ holds iff the set of individuals i for which e_i is the same element as f_i in E_i is a member of D . It can be shown that \approx_D is an equivalence relation. $\prod M_i/D$ is then obtained from $\prod M_i$ by 'filtering through' the equivalence classes generated by this relation.

My key proposal is that *we take the social model structure to be this product*, i.e., we put $M_S = \prod M_i/D$. The social meaning m_S from L to M_S is then defined in the obvious way. The main justification for the equation $M_S = \prod M_i/D$ comes from Los' theorem, which says that if D is an ultrafilter, then a sentence a is true in $\prod M_i/D$ if and only if a is true for a set which is an element of D , i.e., a decisive set of individuals. In the case when D is only a filter, but not an ultrafilter, this equivalence holds only for negation-free sentences. (However, it seems to me that most issues concerning the meaning of expression do not involve negation, so this limitation may be rather harmless). These results justify the claim that the reduced product (or the ultraproduct) is the *emergent* social meaning (given the power structure D and the individual meanings m_i).

If U is finite, all filters are oligarchial, and an ultrafilter is always generated by a single individual i , who then is the semantic dictator. In this case the ultraproduct $\prod M_i/D$ will be isomorphic with m_i . If U is infinite and D is a non-principal ultrafilter, then $\prod M_i/D$ need not be identical with the model of any single speaker. In particular if D is the Fréchet filter or an ultrafilter containing this filter, then no small (i.e., finite) group of individuals will be decisive so that the product $\prod M_i/D$ will be a kind of 'average' of the individual models.

The upshot of my proposal is that the set of individual meaning structures m_i together with the semantic power structure represented by a filter D *uniquely determines* a social meaning in the form of the product $\prod M_i/D$.

¹⁹For an introduction to the theory of reduced products and ultraproducts, see, e.g., Bell and Slomson (1969).

As mentioned above, the models used here are simplistic toy versions and I do not claim that they mirror the semantic complexities of a natural language. One feature that needs to be added is that the meaning of different sublanguages may be dependent on different power structures as was brought up in the previous section. This means that the product construction will be more complicated than depicted here since different sublanguages will be associated with different filters.

Furthermore, I do not believe that the Tarskian type of models is the most appropriate kind, but they should be replaced by more cognitively oriented semantic schemas. For one thing, it seems difficult to drive a wedge between 'truth' and 'meaning' on the basis of such models (I am fully aware that some see this as an advantage of the models). Nevertheless, it seems reasonable that the general strategy followed above – identifying the semantic power structures and applying these in product constructions of individual models – should be applicable also for more authentic models.

The construction outlined here establishes the *existence* of a social meaning, given certain idealizing assumptions. But it does not tell us how a single speaker *determines* what the product M_S (for a given sublanguage) looks like. The solution to this problem depends on the characteristics of the underlying semantic power structure. If the mastery of meaning is dictatorial or oligarchial, the speaker should consult dictionaries, encyclopedias, or other tokens of linguistic power to determine the 'correct' meaning of an expression.

On the other hand, if the power structure is democratic, the right thing to do is to sample assent or dissent from fellow language users. The 'correct' meaning in this case is the answer given by most people. It is intriguing to recall that in the '40s and '50s Arne Næss interviewed ordinary people in Oslo about the meaning of various expressions, e.g., 'truth', as a part of his studies in 'empirical semantics'.²⁰ If the meanings of these expressions were decided by a democratic power structure, this might have been a sound method of investigation; otherwise not.

²⁰See e.g. Næss (1953).

8. The emergent structure as a basis for linguistic convention

Using a language is playing a *game*. We make successful moves in the game when we *coordinate* our expressions so that when I assign a meaning to an expression *a* that I utter (or write) you assign the same meaning to *a* when you hear (or read) it.²¹ If we all were Humpty Dumpties, I would never know whether you mean the same thing by *a* as I do.

In games in general and language in particular the players want to coordinate their actions so that they reach an *equilibrium point*. For those not familiar with game theory, an equilibrium is a set of strategies, one for each individual in the game, in which an individual does as well as he can *given* the strategies of the others. In the present context, a strategy is a way of assigning meanings to the elements of L, i.e., an individual semantics. A situation where the set of individual semantic mappings are *identical* would be an equilibrium since then the speakers would achieve maximal degree of communication (however, there may be other equilibria as well).

The problem is that in a coordination game like language there are in general a large number of equilibrium points (cf. Lewis (1969)). This is true even if we assume that the speakers use the same set L of linguistic expressions. These equilibria correspond to different possible semantic mappings, i.e., ways of assigning meanings to expressions in L. For example various dialects of a language may assign different meanings to the same word. Each such equilibrium is a potential basis for a semantic *convention*. For games with several equilibria we need some procedure for selecting an appropriate equilibrium. This is the *coordination problem* faced by the speakers of a society.

If one adopts the kind of construction of a social meaning outlined in the previous section in a semantic situation S, then the social mapping m_S that emerges in this way *uniquely* identifies an equilibrium point in S. Thus this mapping would solve the coordination problem in S. Even if the speakers start out with different individual semantic mappings m_i , they will obtain better communicative results if they all adjust to the emerging social mapping m_S . In this way m_S serves as a *regulative ideal* for single speakers in analogy with the virtual governor presented in Section 2. Although the mapping m_S need not exist in the head of any individual, the *rational* thing to do is to behave *as if* m_S existed, i.e., to try to adjust one's semantic mapping to the 'virtual governor' m_S .

The game-theoretic analysis given here also explains why it can be said that the emergent social meaning has *causal power*. If the speakers are rational so that they endeavor to attain optimal communication, and if they can identify the emerging social meaning m_S , then this will cause them to change their behavior, i.e., to adjust their individual semantic mappings. Like the individual AC generators in the earlier analogy, the speakers will 'get into step' with the social meaning (cf. Gärdenfors 1990).

²¹For an account of how coordinating strategies may have been established in the evolution of language, see Hurford (1989).

This point about causal power does not apply to language only, but is true of conventions in general. Successful conventions create equilibrium points, which, once established, tend to be stable.²² For example, the convention of driving on the left hand side of the road in Japan will force me to 'get into step' and drive on the left.

In passing, it should be noted that Lewis (1969, p. 42) says that a convention is a 'regularity in behavior'. But the social meaning *m_s* is not 'behavior'. And I don't believe that we can explain what meaning is, social or individual, solely in terms of behavior. Thus, if we adopt Lewis's definition, social meaning in itself cannot be a convention. However, as I have tried to show above, social meaning can uniquely *determine* a convention to be used by the speakers of a language.

Burge (1979) claims that the social character of meaning precludes a purely *individualistic* approach to the mental. The argument starts from his well known example about the patient who uses the word 'arthritis' in a somewhat peculiar way so that he "reports to his doctor his fear that his arthritis has now lodged in his thigh. The doctor replies by telling him that this cannot be so, since arthritis is specifically an inflammation of joints. Any dictionary could have told him the same" (Burge 1979, p. 77). Then Burge invites us to the thought experiment that there is a society the language of which is the same as ours except that "physicians, lexicographers, and informed laymen apply 'arthritis' not only to arthritis but to various other rheumatoid ailments" (p. 78). If we now assume that the patient in the hypothetical society has had exactly the same physical history and the same mental experiences as in the previous case, then Burge claims that the *contents* of the patient's beliefs would still have been *different* in the two situations. Burge concludes that the mental cannot be reduced to the individual: "Social contexts infects even the distinctively mental features of mentalistic attributions. No man's intentional mental phenomena are insular" (p. 87).

The vital point in Burge's argument is what he means by the 'content' of a belief. The two versions of the arthritis example show that the social setting is an important factor in determining the meaning of 'arthritis'. According to Burge this "shows that a person's thought content is not fixed by what goes on in him, or by what is accessible to him simply by careful reflection" (p. 104). He never gives a definition of what he means by 'content', but he presents an 'uncontentious' sufficient condition for separating the contents of different expressions: "On any systematic theory, differences in the *extensions* – the actual denotation, referent, or application – of counterpart expressions in that-clauses will be semantically represented, and will, in our

²²Cf. Popper (1972, pp. 158-159): "We can, in the main, distinguish between two groups of philosophers. The first consists of those who, like Plato, accept an autonomous third world and look upon it as superhuman and as divine and eternal. The second consists of those who, ... , point out that *language*, and what it 'expresses' and 'communicates' is *man-made*, and who, for this reason, see everything linguistic as a part of the first and second worlds, rejecting any suggestion that there exists a third world. ... I think it is possible to uphold a position which differs from that of both these groups of philosophers: I suggest *that it is possible to accept the reality or (as it may be called) the autonomy of the third world, and at the same time admit that the third world originates as a product of human activity.* ... That the third world is not a fiction but exists 'in reality' will become clear when we consider its tremendous effect on the first world, mediated through the second world."

terms, make for differences in content" (p. 75). Again, like in Putnam's argument, we see that Burge sneaks in the external world, via the notion of 'extension', in the characterization of the content of a belief. And if the extension of an expression is determined by social factors, it is not surprising that he can claim that the mental cannot be reduced to the individual.

Burge also argues that the 'content' of an expression is dependent on social conventions. According to him, this is another reason why there can't be any 'independent' individual meanings:

"... the program presupposes that the notion of an individual's believing or intending something is always "conceptually" independent of the conventional meaning of symbols used to express that something. Insofar as 'conceptually' has any intuitive content, this seems not to be the case. Our subjects's belief or intention contents can be conceived to vary simply by varying conventions in the community around him. The content of individuals' beliefs seems sometimes to depend partly on social conventions in their environment. It is true that our subjects are actually rather abnormal members of their community, at least with respect to their use and understanding of a given word. But normality here is judged against the standards set by communal conventions. So stipulating that the individuals whose mental states are used in conventional meaning be relevantly normal will not avoid the circularity that I have indicated. I see no way to do this." (p. 109)

I do. Social meaning need not be defined in terms of 'normal' at all, but what is decisive is the linguistic power structure. The construction of social meaning presented in the previous section provides a way out of Burge's circularity. And it is 'conceptual', at least in the sense that it does not rely on any notion of 'content' that is tied to an external world, i.e., external to the mental structures of the language users and their power relations.

9. Conclusion

The main problem of this article has been to explain the relation between individual semantic mappings and the social meaning of language. In my opinion, this relation is captured by what has here been called the semantic power structure, together with specifications of individual semantic mapping. There is nothing in this construction that violates a conceptualistic approach to semantics (nor does it violate a realistic semantics).

However, as shown above, Putnam (1975, 1988) and Burge (1979) claim that a conceptualistic approach to semantics, *mentalism* as they call it, is doomed to fail. Putnam's main reason for this malediction is summarized by the slogan "meanings ain't in the head." For example, he claims that he cannot distinguish oaks from elms, but he knows that the meaning of the words 'oak' and 'elm' are different. The constructions presented here suggest that, in one sense, Putnam and Burge are right: The social meaning of a locution is not determined by the mental conceptual structure of a single individual.

But Putnam also claims that, as a consequence of this, meanings must be determined by reference to the external world.²³ I believe that this claim is wrong. My position can be summarized as follows: Meanings are not in the

²³And, as I have argued, this is also implicit in Burge's argument.

head of a single individual, but they *emerge* from the conceptual schemes in the heads of the language users together with the semantic power structure. There is no need for a reference external to conceptual structures. Even if Putnam cannot distinguish oaks from elms, they are distinguished in the emergent social semantics. So when he says that he knows that the meaning of 'elm' and 'oak' are distinct, he knows that the *social* meanings differ. In his individual semantics, however, they are undistinguishable.

The result is that Putnam's and Burge's arguments against mentalism are inconclusive. And, as I argued above, Putnam's argument in favor of determining meanings in terms of references to the environment involves an unwarranted form of realistic essentialism. In brief, a conceptualistic semantics is still viable.

Appendix: Proofs of theorems

Theorem 1: If conditions (A1) - (A4) are satisfied, then the set D of decisive sets of individuals forms a filter.

Proof: I will prove the four clauses (i) - (iv) in the definition of a filter given above one by one. (i) $U \in D$ follows immediately from (A2). (ii) If the empty set was decisive, we could take any pair of expressions a and b and apply the definition, finding that $m_S(a) \vdash m_S(b)$, which would contradict (A3b). (iii) (This is the gist of the proof.) Assume that $X \in D$ and $Y \in D$. We want to show that $X \cap Y \in D$. Let S be a situation with $m_i(a) \vdash m_i(b)$ for all $i \in X \cap Y$; we have to prove that $m_S(a) \vdash m_S(b)$ holds. Let S' be a situation with the same logical relationship between $m_i(a)$ and $m_i(b)$ as in S, for all $i \in U$, and where a third expression c is assigned meaning by the individuals so that $m_i(a) \vdash m_i(c)$ and $m_i(b) \vdash m_i(c)$ for each i in $X-Y$, $m_i(a) \vdash m_i(c)$ and $m_i(c) \vdash m_i(b)$ for each i in $X \cap Y$, and $m_i(c) \vdash m_i(a)$ and $m_i(c) \vdash m_i(b)$ for each i in $Y-X$. Such an assignment is always possible according to (A1). Since X is decisive, and for every i in X we have $m_i(a) \vdash m_i(c)$, we must have $m_{S'}(a) \vdash m_{S'}(c)$. Similarly, $m_{S'}(c) \vdash m_{S'}(b)$ since Y is decisive and for every i in Y it holds that $m_i(c) \vdash m_i(b)$. By applying (A3a) we conclude that $m_{S'}(a) \vdash m_{S'}(b)$. Finally, condition (A4) yields $m_S(a) \vdash m_S(b)$. (iv) The condition that if $X \in D$ and $X \subseteq Y$, then $Y \in D$ needs no special conditions for its verification since it follows directly from the definition of decisiveness. *Q.E.D.*

Theorem 2: Assume that U is finite. If the semantic mappings in a situation S satisfies (B1) - (B3), then there is a semantic arbiter in S.

Proof: Suppose that U is finite and that S satisfies (B1) - (B3). We want to show that for all S there exists an $i \in U$ such that for all $a \in L$ $m_S(a) = m_i(a)$. Assume for contradiction that for some S there are expressions a_i such that $m_i(a_i) \neq m_S(a_i)$ for all $i \in U$. By (B3) there are in L names a_i^S of the social meanings $m_S(a_i)$ such that $m_S(a_i) = m_i(a_i^S)$ for all i . This means that $m_i(a_i) \neq m_i(a_i^S)$ for all i . It holds in general of Boolean algebras that $b \neq c$ iff $b \leftrightarrow c \neq$

τ (where $b \leftrightarrow c$ is defined as $(b^* \cup c) \cap (b \cup c^*)$). Thus $m_i(a_i) \leftrightarrow m_i(a_i^S) \neq \tau$ for all i . By compositionality, i.e., (B1), it follows that $m_i(a_i \leftrightarrow a_i^S) \neq \tau$ for all i . Since U is finite we can form the conjunction $\&_{i \in U}(a_i \leftrightarrow a_i^S)$. By general properties of Boolean algebras, we then have, for all i , that $m_i(\&_{i \in U}(a_i \leftrightarrow a_i^S)) \neq \tau$. However, since $m_S(a_i) = m_S(a_i^S)$ for all i , it follows by (B1) again that $m_S(a_i \leftrightarrow a_i^S) = \tau$, for all i , and hence $m_S(\&_{i \in U}(a_i \leftrightarrow a_i^S)) = \tau$. But this contradicts (B2). *Q.E.D.*

Acknowledgements:

The author wants to thank Johan van Benthem, Ingar Brinck, Jens Erik Fenstad, Jan Hartman, Paul Hemeren, Kenneth Holmqvist, Dunja Jutronic-Tihomirovic, Isaac Levi, Rita Nolan, Ryszard Wójcicki, two anonymous referees, and the editors of *Linguistics and Philosophy* for helpful comments and criticism. An early version of this paper was presented at the Seventh Amsterdam Colloquium in Semantics, December 1989, and I am grateful for helpful discussions with the participants. Research for this article has been supported by the Swedish Council for Research in the Humanities and Social Sciences.

References:

- Bell, J. L. and Slomson, A. B. (1969): *Models and Ultraproducts*, North-Holland, Amsterdam.
- van Benthem, J. (1986): *Essays in Logical Semantics*, Dordrecht: Reidel.
- Burge, T. (1979): "Individualism and the mental", *Midwest Studies in Philosophy, vol. 4: Studies in Metaphysics*, Minneapolis: University of Minnesota Press, pp. 73-121.
- Carroll, L. (1871): *Through the Looking-Glass*, Macmillan, London.
- Dewan, E. M. (1976): "Consciousness as an emergent causal agent in the context of control system theory," in *Consciousness and the Brain*, ed. G. Globus, G. Maxwell and I. Savodnik, New York: Plenum Press.
- Fodor, J. A. (1981): *Representations*, MIT Press, Cambridge, MA.
- Gärdenfors, P. (1990): "Social intentions and irreducibly social goods", in *Rationality, Individualism, and Public Policy* (ed. by G. Brennan and C. Walsh), Centre for Research on Federal Financial Relations, Canberra, pp. 91-96.
- Gärdenfors (to appear): "Conceptual spaces as a basis for cognitive semantics", to appear in the proceedings of *Second International Colloquium on Cognitive Science*, Universidad del Pais Vasco, 1991.
- Hansson, B. (1976): "The existence of group preference functions," *Public Choice*, pp. 89-98.
- Hurford, J. R. (1989): "Biological evolution of the Saussurean sign as a component of the language acquisition device", *Lingua* 77, pp. 187-222.
- Jackendoff, R. (1983): *Semantics and Cognition*, MIT Press, Cambridge, MA.
- Jackendoff, R. (1988): "Conceptual semantics," in *Meaning and Mental Representation*, ed. by U. Eco, M. Santambrogio, and P. Violi, Indiana University Press, Bloomington, IN, pp. 81-97.
- Johnson-Laird, P. N. (1983): *Mental Models*, Cambridge University Press, Cambridge.
- Lakoff, G. (1987): *Women, Fire, and Dangerous Things*, University of Chicago Press, Chicago, IL.
- Langacker, R. W. (1986a): "An introduction to cognitive grammar," *Cognitive Science* 10, pp. 1-40.
- Langacker, R. W. (1986b): *Foundations of Cognitive Grammar (Vol. 1)*, Stanford University Press, Stanford, CA.
- Lewis, D. K. (1969): *Convention*, Boston: Harvard University Press.
- Lewis, D. K. (1970): "General semantics," *Synthese* 22, pp. 18-67.

- Lewis, D. K. (1975): "Languages and language," in *Minnesota Studies in the Philosophy of Science, Vol. VII*, ed. by K. Gunderson, University of Minnesota Press, pp. 3-35.
- Næss, A. (1953) *An Empirical Study of the Expressions "True", "Perfectly Certain" and "Extremely Probable"*, Oslo.
- Popper, K. R. (1972): *Objective Knowledge: An Evolutionary Approach*, Oxford University Press, Oxford.
- Putnam, H. (1975): "The meaning of 'meaning'," in *Mind, Language and Reality*, Cambridge University Press, Cambridge.
- Putnam, H. (1988): *Representation and Reality*, MIT Press, Cambridge, MA.
- Radford, R. A. (1945): "The economic organisation of a P.O.W. Camp," *Economica* 12 , pp. 189-201.
- Talmy, L. (1988): "Force dynamics in language and cognition," *Cognitive Science* 12, pp. 49-100.
- Wiener, N. (1961): *Cybernetics*, MIT Press, Cambridge, MA.