Chapter 7

Semantics

Generally speaking, *semantics* is the study of the relation between a *language* and the *environment* in which the language is used. The language can be either *artificial* or *natural*. The former usually has well-defined grammar rules followed by the users of the language (so they are often also called "formal languages" or "symbolic languages"), while the latter is usually formed in history, described by some loose grammar rules summarized from its common usage, which may change from time to time, from place to place, and often with exceptions.

A reasoning system normally uses a language for communicating with other systems, as well as for knowledge representation within the system. For each of the two functions, the semantic theory of the language plays an important role:

- Outside the system, semantics specifies how the language should be understood (such as how to be translated into other languages) in communication, so that other (human or computer) systems know how to "talk" with the system. For this purpose, the semantic theory specifies how the *meaning* of a word or a sentence in the language is determined, by relating it to the outside of the language.

- Within the system, semantics provides justification for the inference rules, that is, to explain why these rules and not others are proper to be used to carry out inference on the language. For this purpose, the semantic theory specifies how the *truth values* of declarative sentences of the language are determined, so that the rules can be validated as preserving truth in the inference process.

In this chapter, the Experience-Grounded Semantics (EGS) of NARS (first introduced in Chapter 3) is compared with other schools of semantics, and several important issues are discussed.

7.1 Experience vs. model

7.1.1 Model-theoretic semantics

The "native language" of NARS is Narsese. It is a formal language, in the sense that its grammar is formally defined (see Table 5.17 for the complete Narsese grammar). Since the dominant semantics for formal languages is Model-Theoretic Semantics (MTS), the first issue to be discussed is why Narsese does not use MTS, but uses a new semantic theory, EGS.

The basic of MTS can be roughly described as the following. For a formal language **L**, a *model* **M** consists of descriptions about objects and their factual relations in a domain. The descriptions are written in a *meta-language* **Lm**, which can either be a natural language, like English, or another formal language. An *interpretation* **I** maps the words in **L** onto the objects and relations in **M**.

According to this theory, the *meaning* of a word in **L** is defined as its image in **M** under **I**, and whether a statement in **L** is *true* is determined by whether it is mapped by **I** onto a fact in **M**.

The study of formal languages was started as part of the study about the foundation of mathematics by Frege, Russell, Hilbert, and others. A major motivation of using formal languages was to avoid the ambiguity in natural language, so that objective and accurate artificial languages were created. MTS was founded by Tarski's work. Although Tarski's primary target was formal language, he also hoped that the ideas could be applied to reform natural language [Tarski, 1944].

To directly use this kind of semantics in a reasoning system (such as NARS) means to understand the meaning of a word in Narsese according to the object or relation it refers to (under a given interpretation), and to choose inference rules that are truth-preserving under all possible interpretations. According to this viewpoint, "semantics is a discipline which deals with certain relations between expressions of a language and the objects 'referred to' by those expressions." [Tarski, 1944].

According to MTS, for any formal language **L**, the necessary and sufficient condition for its terms to have meaning and for its statements to have truth value is the existence of a model. In different models, the meaning of a term and the truth values of a statement may change; however, these changes are not caused by *using* the language. A reasoning system **R** that processes sentences in **L** does not depend on the semantics of **L** when the system runs. That means, on the one hand, that **R** needs no *access* to the meanings of terms and truth values of statements — it can distinguish terms only by their forms, and derive statements from other statements only according to its (syntactically defined) inference rules, but it puts little constraint on how the language can be interpreted [Putnam, 1981]. On the other hand, what beliefs **R** has and what operations **R** performs have *no influence* on the meanings and truth values of the terms and sentences involved.

Such a treatment is desired in mathematics. As Russell put it, "*If* our hypothesis is about *anything*, and not about some one or more particular things, then our deductions constitute mathematics. Thus mathematics may be defined as the subject in which we never know what we are talking about, nor whether what we are saying is true." [Russell, 1901] In mathematical logic, abstract patterns of inference are studied, and the patterns can be applied to different domains by constructing different models. Here we do enjoy the freedom provided by the separation of "syntactic processing" and "semantic interpretation." The study of semantics has contributed significantly to the development of meta-mathematics. As Tarski said, "As regards the applicability of semantics to mathematical science and their methodology, i.e., to meta-mathematics, we are in a much more favorable position than in the case of empirical sciences." [Tarski, 1944].

As all normative theories, MTS is based on certain assumptions, and it should be applied to a problem only when the assumptions are satisfied. In asserting the existence of a model **M**, the theory presumes that there is, at least in principle, a consistent, complete, accurate, and static description of (the relevant part of) the environment in a language **Lm**, and that such a description, a "state of affairs" is at least partially known, so that the truth value of some statements in **L** can be determined accordingly. These statements then can be used as premises for the following inference activities. It is also required that all valid inference rules must be truth-preserving, which implies that only true conclusions are desired. After the truth value of a statement is determined, it will not be influenced by the system's inference activity.

Such conditions hold only when a system has *sufficient knowledge and resources* with respect to the problems to be solved. "Sufficient knowledge" means that the desired results can be obtained by deduction from initially available knowledge alone, so no additional knowledge will be necessary; "sufficient resources" means that the system can afford the time–space expense of the inference, so no approximation will be necessary. These are exactly the assumptions we usually accept when working within a mathematical theory. Therefore, it is no surprise that model-theoretic semantics works fine there.

Of course, what we just described is merely the basic form of MTS. Many variations and extensions of MTS have been proposed for various purposes, such as possible worlds, multi-valued propositions, situational calculus, and so on [Barwise and Perry, 1983, Carnap, 1950, Halpern, 1990, Kyburg, 1992, Zadeh, 1986b]. However, these approaches still share the same fundamental framework: for a reasoning system **R** working in an environment **E** with a language **L** (for knowledge representation and communication), the semantics of **L** is provided by descriptions of **E** in another language **Lm** and a mapping between items in **L** and **Lm**.

No matter how the details are specified, this kind of semantics treats the semantics of **L** as *independent* of the two processes in which **R** is involved (and where the language **L** plays a central role): first, the communication between **R** and **E**, and second, the internal reasoning activity of **R**. According to MTS, these processes are purely syntactic, in the sense that only the form of the words and the structure of the sentences are needed. Since the above two processes can be referred to as the "external experience" and "internal experience" of the system, we say that MTS is "experience-independent," and it does not even need to assume the existence of a reasoning system **R** that actually uses the language.

7.1.2 Why NARS does not use MTS

Though MTS can be applied to Narsese, it provides little help for the design and use of NARS. If we give Narsese a model, it tells us what the words mean to *us*, but says nothing about what they mean to *the system*, which does not necessarily have access to our model. Similarly, the model tells the truth value of statements to *us*, but not to *the system*.

By "to the system," I mean that to solve the semantic problems in NARS (that is, to understand the language and to justify the rules), we need to explain why the system treats each term and statement as different from other terms and statements, and such an explanation should be based on the relation between the language and the environment, not only on the syntactic natures defined within the language. Since the relation between NARS (the system in which the language is used) and its environment (which is the "world" to the system) is indicated by the *experience* of the system, the semantic features of a term or a statement has to be determined according to its role in the experience of the system, because in NARS there is no other way to talk about the outside world.

If we still define truth as "agreement with reality," in the sense that truth values cannot be threatened by the acquisition of new knowledge or the operation of the system, then no statement can ever be assigned a truth value by the system under AIKR, because by the very definition of *open* system, all beliefs can be challenged by future evidence.[1]

Moreover, since non-deductive inferences (which are absolutely necessary when knowledge is insufficient) are not truth-preserving in the model-theoretic sense, they are hardly justifiable in the usual MTS way. MTS also prohibits the system from using the same term to mean different things in different moments (which is often inevitable when resources are in short supply, to be discussed later), because meaning is defined as independent of the system's activity.

However, it is not true that in such a situation semantic notions like "truth" and "meaning" are meaningless. If that were the case,

[1]One proposed solution of this problem is to treat "truth value" and "degree of belief" as different to each other. This issue will be discussed in detail in Section 7.4.2.

then we could not talk about truth and meaning in any realm except mathematics, because our mind faces exactly the same situation.

For an intelligent system likes NARS (or for adaptive systems in general), the concept of "truth" still makes sense, because the system *believes* certain statements, but not other statements, in the sense that the system chooses its actions according to the expectation that the former, not the latter, will be confirmed by future experience; the concept of "meaning" still makes sense, because the system uses the terms in Narsese in different ways, not because they have different shapes, but because they correspond to different experiences.

For these reasons, in NARS we need an *experience-grounded* semantics, in which truth and meaning are defined according to the experience of the system. Such a theory is fundamentally different from MTS, but it still qualifies to be a "semantics," in a broad (and original) sense of the notion.

The idea that truth and meaning can be defined in terms of experience is not a new one. For example, it is obviously related to the theory of pragmatism of Peirce, James, and Dewey. In recent years, related philosophical ideas and discussions can be found in the work of Putnam and many others [Dummett, 1978, Field, 2001, Fodor, 1987, Lynch, 1998, Putnam, 1981, Segal, 2000, Wright, 1992]. In linguistics and psychology, similar opinions can be found in [Barsalou, 1999, Ellis, 1993, Kitchener, 1994, Lakoff, 1988, Palmer, 1981].

In AI research, the situation is different. Unlike in philosophy, linguistics, and psychology, where MTS (with the related theories, such as realism, the correspondence theory of truth, and the reference theory of meaning) is seen as one of several candidate approaches in semantics (by both sides of the debates), in AI not only is MTS accepted explicitly by the "logical AI" school [McCarthy, 1988, Nilsson, 1991], and implicitly by the "symbolic AI" school in general [Newell, 1990], but also it is taken to be the only possible semantics, both by its proponents and its critics. As McDermott put it, according to the logicist opinion, "The notation we use must be understandable to those using it and reading it; so it must have a semantics; so it must have a Tarskian semantics, because there is no other candidate." [McDermott, 1987] When people do not like this semantics, they usually abandon it together with the idea of formal language and inference rules, and turn to neural networks, robots, dynamic systems, and so on, with the hope that they can generate meaning and truth from perception and action [Birnbaum, 1991, Brooks, 1991, Harnad, 1990, Smolensky, 1988, van Gelder, 1997].

Therefore, though the philosophical foundation of MTS is under debate, and its suitability for a natural language is doubtable [Haack, 1978], few people have doubt about its suitability for a formal language. We have not seen a *formal semantics* that is not *model-theoretic*, and such concept may even sound self-contradictory to some people.

7.1.3 EGS vs. MTS

The EGS theory used in NARS has been formally defined in Chapter 3. Briefly speaking, an EGS first defines the form of experience a system can have, then defines truth value and meaning as functions of given experience.

Though both are descriptions of an environment (or "world"), "model" and "experience" are different in the following aspects:

- A model is static, whereas experience stretches out over time.
- A model is a complete description of (the relevant part of) an environment, whereas experience is only a partial description of it, in the sense that novel terms may appear that were not known previously.
- A model must be consistent, whereas judgments in experience may conflict with one another.
- A model of a language **L** is described in another language **Lm**, whereas experience is represented in **L** itself.
- The existence of a model **M** of **L** is independent of the existence of a system **R** using **L**. Even when both **M** and **R** exist, they are not necessarily related to each other in any way. On the contrary, experience must happen in a system.

These two types of descriptions serve different purposes. A reasoning system assuming sufficient knowledge and resources makes no attempt to answer questions beyond the scope of available knowledge and resources — when such a question is provided, the system simply replies "I don't know," "Invalid question," or gives no reply at all. For such a system, it is fine to describe its environment as a model. On the contrary, a system designed under AIKR always attempts to answer a question with available knowledge and resources, which means that the system may revise its beliefs from time to time. For such a system, it is better to describe its environment by its experience.

Generally speaking, the human mind works with insufficient knowledge and resources [Medin and Ross, 1992]. However, for certain relatively mature and stable beliefs, it is more efficient to treat them assuming the sufficiency of knowledge and resources. This is exactly the role played by mathematics. In such a theory, we do not talk about concrete objects and properties. Instead, we talk about abstract ones, which are fully specified by postulates and conventions. After we figure out the implications of these postulates and conventions, we can apply such a theory into many situations, because as far as the postulates and conventions can be "instantiated" by substituting the abstract concepts with the concrete ones, all the ready-made implications follow. This is the picture provided by MTS.

On the other hand, if the beliefs embedded in a reasoning system are not mathematical, but empirical, then what we have is a system where the concepts are no longer abstract and can be interpreted freely — no matter how an external observer interprets them, for the system their meaning and truth come from experience, and an EGS should be used.

Some researchers suggest that the reasoning system itself (human or computer), rather than the world it deals with, should be used as the "domain" of the language the system uses. Thus, one could posit that the meaning of a particular term is a particular "concept" that the system has, and the truth value of a statement is the system's "degree of belief" in that statement. This idea sounds reasonable, but it does not answer the original question: how are "concepts" and "degrees of belief" dependent upon the outside world? Without an answer to that question, such a solution "simply pushes the problem of external significance from expressions to ideas" [Barwise and Perry, 1983], that is, it turns the problem of *word* meaning into the problem of *concept* meaning.

The meaning of a concept is not simpler than the meaning of a word at all. It often changes from time to time and from place to place, and such changes cannot always be attributed to the changes in the world. People in different cultures and with different languages usually have different opinions on what "objects" are there even if they are in the same environment [Whorf, 1956]. People often use concepts metaphorically [Lakoff, 1987] or with great "fluidity" [Hofstadter and FARG, 19995]. These issues are hard to handle in MTS.

What if we take a concept as a Platonic entity that never changes, and treat the changes in the meaning of a word as mappings to different concepts? Now the problem becomes to explain why a certain new concept, rather than many others, becomes the new meaning of a word. We still need a way to link the change of meanings of words, concepts, ideas, or whatever we call them, to the experience of the system.

In NARS, since a term is the name of a concept, and a statement is the name of a conceptual relation, the meaning/truth defined for the language and the meaning/truth defined for the concepts system are defined similarly.

Though overall NARS uses EGS, there are still places where MTS is used. One example is the symbols in the inference rule. As mentioned in Chapter 3, the induction rule of NARS is

$$\{M \rightarrow P < f_1, c_1 >,\ M \rightarrow S < f_2, c_2 >\} \vdash S \rightarrow P < f, c >$$

Written in this way, the symbols S, M, and P have no experience-related meaning until they are instantiated by constant terms "*bird*," "*raven*," and "[*black*]," respectively, and then the meanings of the symbols are determined by the meanings of the constant terms.

Similarly, when mathematical knowledge is provided to NARS, it is used with MTS. This kind of knowledge always needs an interpretation step when being applied to a practical situation. Roughly speaking, MTS is usually used to base one language in another language, while EGS is usually used to base a language in the experience of a system using the language.

According to the above discussion, we see that MTS and EGS are designed under different assumptions, and therefore should be used for different purposes. In this sense, EGS is not proposed as a competitor of MTS. However, since now MTS is used everywhere, including in situations where EGS (or its variations) should be used, EGS indeed competes with MTS as candidates of application, especially in AI and cognitive sciences (but not in meta-mathematics).

It is very often implicitly assumed that the semantics of a formal language has to be model-theoretic. Such an inductive conclusion seems warranted by our experience — almost all formal languages have traditionally been assigned their semantics in this way. As a result, people who do not like the semantics usually abandon the language at the same time.

However, a language can be "formal" in two different senses. In a *syntactic* sense, "formal" means merely that the language is artificial, and is defined by a formal and symbolic grammar; in a *semantic* sense, "formal" means that the language is used in conjunction with an MTS. Narsese is "formal" in the syntactic sense only. From a technical point of view, it would be easy to give the language a model-theoretic semantics, but with such a semantics, the language would no longer be suitable for our current purposes.

Logicians, in distinguishing themselves from other scholars (such as psychologists), tend to stress the *normative* nature of logical theory. As a result, in their study of semantics, the goal is often that of looking for *the* real, objective meanings of terms or truth values of sentences. Even if such an opinion has some degree of justifiability when one's purpose is to study the logic of mathematics, that justifiability goes away when one turns to the study of the "logic" of empirical science and common sense. For the purposes of AI, what we need is another kind of normative model, in which meanings and truth values are founded on the system's experience.

Since NARS is a *normative* theory of intelligent reasoning, not a *descriptive* theory of it, the semantics proposed here is about how truth and meaning *should be* used in a system, not how they *are* actually used in the human mind. I do not present NARS as a psychological or linguistic model of truth and meaning. However, since the human mind is basically an adaptive system evolved in an environment where its knowledge and resources are generally insufficient with respect to the problems to be solved, I do believe that in general this model is closer to a descriptive model than MTS is. Though it is not the major goal of the current research on EGS, it will be interesting to explore the implications of this theory in philosophy, linguistics, and psychology.

7.2 Extension and intension

Traditionally, *extension* and *intension* refer to two different aspects of the meaning of a term: roughly speaking, its *instances* and its *properties*.

In previous theories, a term's extension is usually defined as a set of *objects* in a "physical world" that are denoted by the given term; a term's intension is usually defined as a *concept*, or a set of *attributes*, in a "Platonic world" which denotes or describes the given term [Bocheński, 1970, Copi, 1982]. In spite of minor differences among the exact ways the two words are used by different authors, they always indicate relations between a term in a language and something *outside* the language.

By contrast, in NAL (as defined in Chapter 3) a term's extension and intension are sets of term linked to the given term by an inheritance relation (in the opposite directions). Here extension and intension are defined *within* the language, and become symmetric to each other. Yet even so, the definition retains the intuitive feature that "extension" refers to instances, and "intension" refers to properties.

Such a departure from tradition has important reasons and implications. One of them is already addressed previously in the "experience vs. model" discussion, that is, NAL does not assume a "physical world" or a "Platonic world" that is described in another language, and everything that is semantically relevant must be based on the experience of the system, described by the same language. In the following, we focus on the symmetry, or duality, of extension and intension in NAL, which lead to a unified treatment of the two.

7.2.1 The need for a unification

One feature that distinguishes NAL from other logical systems is its unified representation and processing of extension and intension.

In semantics, this unification happens in two places. For each term, its meaning consists of its extension and intension. For each judgment, its truth value is usually determined by both extensional and intensional factors.

It needs to be stressed that in the terminology of set theory, what is being counted in NAL as a piece of evidence in extension is not an "element," but a "subset," because of the use of inheritance relation. For example, if the system's experience is $\{A \circ\!\!\rightarrow B,\ B \rightarrow C,\ C \rightarrow D\}$, then the extension of D is $\{\{A\}, B, C\}$, not just $\{A\}$. Similarly, what is being counted in intension is a "superset," not a "property."

In other theories, following the tradition of set theory, a concept is often treated as a set, with its instances as elements. Consequently, the logic developed on these concepts is a kind of extensional logic. If intension is as important as extension, how can the traditional logic have been used for such a long time without being challenged on this issue?

This is the case because traditional formal logic has been developed and mainly used in mathematics, where given the extension of a concept (i.e., what instances it has), its intension (i.e., what properties it has) is uniquely determined; and given its intension, its extension is uniquely determined, too. We have seen this in the discussion of NAL-0 in Section 3.1. In such a situation, to process both extension and intension becomes unnecessary, and even confusing. When concepts are explicitly defined and processed according to their extension, their intension is implicitly defined and processed.

This is no longer the case for a system like NARS, which is built under AIKR. In this situation, not only that extension and intension do not fully determine each other, even known extension (intension) cannot determine future extension (intension).

For example, in set theory, $S \subseteq P$ means that the members of S are also members of P. This extensional statement implies the following:

- if M is an element of S, it is also an element of P;
- if M is a property shared by the elements of P, it is also shared by the elements of S.

On the contrary, in NAL if all known instances of S are also instances of P, the system is not necessarily certain to the same extent (i.e., indicated by the same truth value) about the implications listed above. What it means is that in this situation, we cannot only process extension of concepts, and expect intension to follow automatically.

For practical purposes, we may prefer to treat concepts as purely defined by extension (or intension), so that we can use various mathematical tools on them. By doing that, however, we are assuming certain beliefs to be "axioms" that won't be challenged, and doing inference accordingly. In that situation, it is fine to concentrate on extension alone. For example, we can define a "subset" relation among concepts (that are treated as sets), and do inference on the relation accordingly. The pure extensional logic used in this case is not part of NAL, but a system that can be called by NAL as a tool to solve specific problems. NAL is not designed as a logic that *includes* all other useful logics, but one that allows the others to be used by it. To a system using NARS as its "intelligent core," it works like an operating system, which use various logics and algorithms as application programs to solve various concrete problems.

However, no matter what logic is used, the conclusions obtained are only as good as the assumptions, and the price of ignoring intensional information will be paid anyway. If certain concepts should not be treated as pure extensional, but the system chooses to do that anyway, the conclusions will be less confident than the ones obtained by considering both extensional and intensional beliefs, because the former is based on less evidence.

It is possible to develop extensional or intensional term logics separately, as shown in [Wang, 1994b]. As stated previously, "$S \rightarrow P$" means, when it is understood *extensionally*, that P inherits S's instances; but when it is understood *intensionally*, the same relation means that S inherits P's properties. Therefore, if "$S \rightarrow M$" is completely false and "$M \rightarrow P$" is completely true, what can be derived from them is different in the two logics. In the extensional logic, the premises are understood as "S and M have no common instances, and all instances of M are also instances of P." From these two relations, we cannot decide whether S and P have common instances. On the other hand, in the intensional logic, the premises are understood as "S and M have no common properties, and all properties of P are also properties of M," which implies that "S and P have no common properties."

Symmetrically, if "$S \rightarrow M$" is completely true and "$M \rightarrow P$" is completely false, the extensional implication is "S and P have no common instances," and there is no intensional implication.

Though the extensional logic and the intensional logic, defined in this way, are different, formally they are isomorphic to each other, much as union and intersection are isomorphic to each other in set theory. This isomorphism is described in [Wang, 1994b], and it comes directly from the "dual" definitions of extension and intension in NARS.

The dual definitions of extension and intension make it possible for NARS to treat them *uniformly*, as, for instance, in the definition of "amount of evidence." We need systems to deal with them together, because the coordination of the extensions and intensions of concepts is an important principle in the development of human cognition [Inhelder and Piaget, 1969], and when evidence is used to judge a conceptual relation, whether the evidence is extensional or intensional is often irrelevant or unimportant. We often determine the extension (instances) of a concept according to its intension (properties), or the other way around, and seldom judge a relation between concepts by considering the extensional or intensional factor *only*, especially when the system has insufficient knowledge and resources.[2]

Therefore, though a pure extensional (or intensional) logic is easier to define, it is less interesting from the viewpoint of AI, so NARS is not designed in that way. On the other hand, if really necessary, it is possible for NARS to express pure extensional or pure intensional relations using more complicated methods (such as with variable terms defined in NAL-6), and to support an extensional (or intensional) logic as a subsystem.

7.2.2 Unification in meaning

As discussed above, the meaning of a term has two aspects, its extension and intension, related to the "reference" and "sense" of Frege, respectively [Copi, 1982].

A common practice in AI and cognitive science is to take a term as the name (or label, symbol, and so on) of an object, or a set of objects, in the world. This intuition is the foundation of the MTS.

[2]We will come back to this issue in Chapter 11 when discussing categorization.

The problem of this approach is the assumption that there is an objective way to describe the world as objects with relations among them, and that terms have one-to-one mapping to objects or sets of object. This assumption, though sounds natural, conflicts with the assumption that the system may have all possible kinds of future experience.

As usual, this idea is not completely wrong. The meaning of a term does partially depend on its relation to its instances (extension). The difference is that in NARS the extension of a term consists of other *terms*, not *objects*. By following the extensional relations, the system will eventually reach terms that cannot be further specified, though actually they still exist within the system, for instance, as the "mental image" that formed by perception. For example, we can say that "Mars" is in the extension of "planet," but here "Mars" is something in our mind, not something that exists independent of our mind.

According to EGS, the meaning of a term is determined by its (experienced) relations with other terms. These relations can either be extensional or intensional. If "$S \rightarrow P$" is a new belief, then it contributes to the meaning of both S (by indicating part of its intension) and P (by indicating part of its extension). In a sense, S and P are partially defining each other, and no one is "more primitive" than the other semantically.

As special cases, there are terms whose meaning is mostly determined by its extension, as well as terms whose meaning is mostly determined by its intension. Only in these situations, a pure extensional or pure intensional theory of categorization may work approximately. Since NARS determines the meaning of a term according to *available* relations, it can handle these special situations, as well as the general situation where both extensional and intensional factors should be considered.

One consequence of using a unified approach is that that system tends to keep the extension and intension of a concept in coherence. In an extensional logic, since properties are derived from given instances, the system makes no attempts to use them to evaluate membership. It is possible to define a concept by a set of instances, where one instance is very different from the others (in terms of its properties), and yet is a perfect instance. In NARS, however, the result is different. The membership (i.e., the frequency) of the special instance will be decreased, because of its difference with the other instances. In this way, there is a "feedback loop" between extension and intension.

7.2.3 Unification in truth value

According to the definitions introduced in Chapter 3, in the truth value of a judgment, the extensional factor and the intensional factor are merged together.

This is a controversial issue in NARS. Intuitively, even if both extensional and intensional factors need to be considered, it is more informative to represent them separately, and to process them in parallel. For example, for "$S \rightarrow P$," why do not we measure the evidence collected from the extension and intention of S and P with different numbers? As shown in [Wang, 1994b], it can be done in binary logic, and the same idea could had been applied to NAL.

NARS does not follow that path because, once again, AIKR. Here this assumption means "to make judgment according to whatever knowledge is available," no matter where it comes from.

Let's take medical diagnosis as an example. Suppose a doctor wants to determine whether a patient P is suffering from disease D, that is, to evaluate statement "$\{P\} \rightarrow D$." For this task, at least two types of information can be taken into account: (1) whether P has D's symptoms S (that is, to derive the conclusion from "$D \rightarrow S$" and "$\{P\} \rightarrow S$" by abduction), and (2) whether D is a common illness among reference class C to which P belongs (that is, to derive the conclusion from "$C \rightarrow D$" and "$\{P\} \rightarrow C$" by deduction). With respect to the term D, the inference is intensional in (1), and extensional in (2). To get a summarized conclusion means to merge the two conclusions by the revision rule, and the result is neither pure extensional nor pure intensional for D. Such a merging also means that different types of evidence (extensional and intensional) can be balanced against each other, or be accumulated together.

After the truth value of "P is suffering from D" is evaluated, it can be combined with the truth value of "T is a proper treatment to D" (which is usually a statistic statement, too, therefore extensional) to get the truth value for "T should be applied to P." In such a situation (which is the usual case, rather than an exception), even if extensional and intensional evidence can be distinguished in the premises, they are mixed in the intermediate and final conclusions. If the system insists in separating extensional and intensional truth values, the above inference cannot be carried out.

Here we get to this conclusion once again, though along a different path: technically, it is possible to build a pure extensional or intensional logic, but when they are used with insufficient knowledge and resources, there are many situations where they cannot use the available knowledge like NAL does. Compared to NAL, they are not "wrong," but "weak."

If extensional and intensional evidence are collected in different ways, is it valid to merge them into a single truth value? It is valid, because though they are from different sources, the evidence contributes in the same way to the truth value. In the design of truth-value functions, no assumption is made on whether the evidence of the premises are extensional or intensional, so the system is consistent on this issue.

Sometimes we do hope to distinguish extensional and intensional factors in truth value for the purpose of explanation, such as in answering the question "Why do you believe that S is a special kind of P?." In this case, "Because S has the properties of P" and "Because P has the instances of S" are obviously different. However, we do not keep this information in the truth value of the belief "$S \rightarrow P$," because the truth value is used to summarize evidence, not to keep detailed information about evidence.

With insufficient knowledge and resources, the system makes no attempt to keep all the information about how a conclusion is obtained in the truth value of a statement. If we really need to separate the extensional factor and the intensional factor of "$S \rightarrow P$," it can be done by instead talking about "$(\#x \rightarrow S) \Rightarrow (\#x \rightarrow P)$" and "$(P \rightarrow \#x) \Rightarrow (S \rightarrow \#x)$," as mentioned previously.

7.3 Meaning of term

7.3.1 Meaning in NARS

The definition of meaning in EGS has the following implications:

1. The meaning of a term is its experienced relations with other terms.

2. The meaning of a term consists of its extension and intension.

3. Each time a term is used in an inference process, only part of its meaning is involved.

4. Meaning changes with time and context.

5. Meaning is subjective, but not arbitrary.

As said previously, a human observer can still interpret the terms appearing in NARS freely by identifying them with words in a natural language or human concepts, but that is their meaning *to the interpreter*, and has little to do with the system itself. For example, if the term "*bird*" never appears in the system's experience, it is meaningless to the system (though meaningful to English speakers). When "$bird \rightarrow animal < 1, 0.8 >$" appears in the system's input stream, the term "*bird*" begins to have meaning to the system, revealed by its inheritance relation with "*animal*." As the system knows more about "*bird*," its meaning becomes richer and more complicated. The term "*bird*" may never mean the same to NARS as to a human (because we cannot expect a computer system to have human experience), but we cannot say that "*bird*" is meaningless to the system for this (human chauvinistic) reason. This is just like that a child often uses a word in a different way, compared to its common usage. We can say that the usage is "different," or even "wrong," but we cannot say that the word is "meaningless" to the child. As long as a term has experienced relations with other terms, it becomes meaningful to the system, no matter how poor its meaning is.

An adaptive system should never processes a term only according to its shape without considering its position in the system's experience. The shape of a term may be more or less arbitrary, but its experienced relations with other terms are not.

This conclusion to an extent agrees with Wittgenstein's claim that the meaning of a word is its use in the language [Wittgenstein, 1999]. For NARS, the meaning of a term, such as "*game*," is not determined by a definition or a set of "things" in the world, but by how the term is related to the other terms according to the system's experience. As a result, there may be no common property shared by all instances of "*game*." Instead, there is only a "family resemblance" among them, indicated by the overlapping properties here or there (without a definitive property for all of them).[3]

7.3.2 Symbol grounding

By saying the above, I do not mean that in the human mind a word in a natural language gets its meaning *only* by its relation with other words in the language, because *human experience* is not limited to a language channel, but closely related to sensation, perception, and action [Barsalou, 1999, Harnad, 1990]. However, the general principle is still applicable here, that is, a word gets its meaning by its experienced relations with the system's other *experiential components*, which may be words, perceptive images, motor sequences, and so on. In a system like this, the meaning of a word is much more complex than in a system whose experience is limited to a language channel, but it does not rule out the latter case as a possible way for words (terms, symbols) to be meaningful. For example, a software agent can get all of its experience in this manner, and we cannot deny that it is genuine experience.

For a symbolic system built according to an EGS, the symbols in the system are already *grounded* — in the system's experience, of course. The crucial point here is that for a symbol to be meaningful (or grounded), it must be related somehow to the environment. However, such a relation is not necessarily via a sensorimotor mechanism. The experience of a system can be *symbolic*, as in the case of NARS. This type of experience is much simpler and "coarse-grained" than sensorimotor experience, but it *is* real experience, so it can ground the symbols which appear in it, just as words in natural language are grounded in human experience. In the future, when NARS can accept visual input, an image will be related to the concept of "Mona Lisa," so it does not merely mean "a painting by Leonardo da Vinci." This additional link changes the meaning of the concept, but it does not change the semantic principle of the system: the meaning of the concept is not completely determined by the "object in the world referred to by it," but by its experienced relations with other things in the system's experience.

[3]In this way, the semantics of NARS also implies a new theory of categorization, which will be discussed in Section 11.1.

The definition of meaning in EGS is similar to that of conceptual role semantics and semantic network [Harman, 1982, Kitchener, 1994, Quillian, 1968], where the meaning of a concept (or word) is defined by the role it plays in a conceptual system (or a natural language). The difference between EGS and these theories are:

- In NARS, the relations among terms are not definitional or conventional, but are *experienced* through the interaction between a system and its environment. Therefore, they are dynamic and subjective in nature.
- In NARS, the relations between a term and others are concretely specified by its extension and intension, consisting of inheritance judgments, whose meaning and properties are formally specified.
- In NARS, whenever a term is used, only part of its meaning is involved. In other words, the "current meaning" of a term is not exactly its "general meaning" in the long run.

Similar ideas are called "dictionary-go-round" by Harnad — he hopes that meaning of symbols can "be grounded in something other than just more meaningless symbols" [Harnad, 1990]. Here we should notice a subtle difference: in EGS, the meaning of a term is not *reduced into (or grounded on) the meaning of other terms* (that will indeed lead to circular definition in a finite language), but *defined by its relations with other terms*. These relations are formed during the interaction between a system and its environment, and are not arbitrary at all.

Another relevant factor is that in NARS, the copulas, term operators, and other syntactic markers are logical constants in Narsese. Their meaning is innate to the system, because they are directly recognized and processed by the inference rules and control mechanism in predetermined ways. Even when all the terms in an input statement are novel, the inheritance relation is known. Therefore, NARS is not "getting meaning out of the meaningless."

For these reasons, though NARS has a language defined by a formal grammar and used by a set of formally defined inference rules, it is not a "symbol system" discussed by Harnad, where symbols get their meanings "as standing for objects, as describing states of affairs" [Harnad, 1990].

7.3.3 Chinese room

This leads us to Searle's "Chinese room" argument [Searle, 1980], by which he claimed that a system using syntactic rules cannot have meaning.

Searle's argument is based on the assumption that a symbol can get meaning only from a model, by an interpretation. If one accepts the idea of an EGS, this is an untenable argument. As said above, as soon as a term has experienced relations with other terms, it becomes meaningful to the system, no matter how impoverished or diluted its meaning is. An adaptive system like NARS never processes a term solely on the basis of its shape, without considering its relations with other terms in the system's experience.

The feeling of meaninglessness in Searle's "Chinese room" comes from his deliberate cutting-off of his experience in Chinese from his sensorimotor experience and his experience represented in his native language. If we put an intelligent computer system into the same situation, there are two possible cases. If the computer system already had profound sensorimotor experience and/or a "native language," it might also consider the Chinese characters to be meaningless, because it could not relate them to its previous experience.

However, if the system entered the room with no previous experience, Chinese would become its "native language" — that is, the system would build up meanings for the characters on the basis of how they are related to one another, and would not attempt to ground them on some "more fundamental" stuff, nor would it complain about "meaningless squiggles and squoggles" when it failed in doing so. If the system also had sensorimotor capacities and communicated with other similar computer systems in Chinese, we might find that the meanings of Chinese words, to these systems, were as rich and as complex as they are to human Chinese speakers, though it is possible that they might occasionally have different opinions about the "correct" meaning of a given word.

Please note that my response to Searle's problem is not "the robot reply" he rejected in his paper [Searle, 1980]. I agree with him that even if a sensorimotor mechanism is introduced into the system, it does not directly bring meanings to symbols. According to EGS, a term is meaningful if it has experienced relations with other components of the system. Here, what matters is whether a symbol has (recognizable) relations in the system's experience, not whether the experience comes from a sensorimotor mechanism.

7.3.4 Subjectivity

As mentioned in Chapter 6, due to insufficient resources, the system cannot consult all known beliefs associated with a term each time the term is used. Instead, in NARS a priority distribution is maintained among the beliefs, which determines the chance for a certain belief to be taken into consideration at the current time. The distribution is adjusted by the system according to the feedback of each inference step (to make the more useful beliefs more accessible), as well as according to the current context (to make the more relevant beliefs more accessible).

Consequently, the meaning of a term becomes context-dependent — it does not only depend on what the system knows about the term, but also depends on the system's current tasks and how the relevant beliefs are ranked in terms of their priority. When the system gets new beliefs, or turns to another task, the meaning of the involved terms may change (more or less). Again, these changes are anything but arbitrary, and the meaning of some terms may remain relatively stable during a certain period. Without such a restriction, a "relational" theory of meaning cannot be practically used, because in a sufficiently complicated system, a concept may (in principle) be related to other concepts in infinite number of relations, and to take all of them into account is impossible.

Since the meaning of a term is determined by the system's experience, it is fundamentally subjective. However, as soon as the term is used in the communication with another system, the two systems begin to have common experience, and they begin to know how the term is used by the other. In the long run, the meaning of such terms gradually become "objective" in the sense that it reflects the common usage of the term within the language community, and less biased by the idiosyncratic usage of a single system.

Therefore, we can still understand what NARS means by a certain term and agree with a belief of the system, because of the partial overlap of its conceptual system with ours. However, we cannot expect its conceptual system to be identical to that of a human being, due to the fundamental difference between its experience and our experience.

Accurately speaking, no two people have identical conceptual systems (so misunderstanding and disagreements happen all the time), but we can still communicate, and understand each other to various extents on various topics, because we co-exist in the same world and the same human society, therefore have shared (physical and social) experience.

7.4 Truth of statement

7.4.1 Truth in NARS

As defined in Chapter 3, in NARS "truth" corresponds to statements with truth value $< 1, 1 >$, and non-analytic truth can only be approached, but not reached, by actual beliefs in the system. In general, in NARS truth is a matter of degree, represented by a $<f, c>$ pair.

The definition of truth value in the semantics of NARS has the following implications:

1. Truth is a matter of degree, and, in the case of first-order statements, is determined by the extent to which the subject term and the predicate term of the statement can be substituted by each other in certain ways.

2. A truth value consists of a pair of real numbers, one for the relative amount of positive evidence (with respect to negative evidence), and the other for the relative amount of all available evidence (with respect to future evidence).

3. A truth value is assigned to a statement according to the *past* experience of the system. It does not indicate whether the statement will be consistent with *future* experience, though an adaptive system behaves according to it.

4. For a statement, in the same time the system may has it in several beliefs, with different truth values, derived from different parts of the system's experience. Which one will be used at a given time depends on many factors, including the priority distribution

among the beliefs. As a result, the truth value of a statement seems to change from context to context.

5. Since truth comes from the experience of the system, it is subjective (i.e., different systems usually have different opinions on what is truth), but not arbitrary (i.e., such an opinion can be explained according to the experience of the system).

It is well known that when the evidential support of a scientific hypothesis is evaluated, we not only pay attention to the *amount* of (positive and negative) evidence, but also to its *diversity*, that is, we hope the evidence to be different in other aspects, while remain to be evidence for the hypothesis under evaluation. In NARS, the diversity of evidence is not directly measured in the amount of evidence (and therefore, in truth value), but works indirectly. When a hypothesis is supported by uniform evidence, the same evidence usually also support some other competing hypotheses, and therefore will eventually make the hypothesis unlikely to be chosen. On the contrary, diverse evidence means that the other factors show less regularity than the one captured in the hypothesis, therefore the competing hypothesis get less support, and the hypothesis under consideration is more likely to be chosen (even though its truth value may remain the same as the case of uniform evidence).

7.4.2 Truth value and degree of belief

EGS is similar to the "coherence" theories of truth [Haack, 1978], in that the truth value of a belief is partially determined by the truth values of other beliefs in the system, and the system will try its best to resolve inconsistency among its beliefs. However, in NARS some of the beliefs come from the experience of the system, so they are not necessarily consistent initially, and the system usually cannot achieve complete consistency among its beliefs, no matter it has tried. On the other hand, the system will not accept a set of consistent beliefs if it is not related to its experience.

On the contrary, MTS provides a "correspondence" theory of truth [Haack, 1978], where the truth value of a statement is determined by whether it corresponds to the state of affairs, as described in a meta-language. According to MTS, "truth value" and "degree of belief" are fundamentally different — a system can strongly believe a false statement. This is from the viewpoint of an observer who knows the "objective truth" and can compare it with a system's belief. However, for the system itself, if it has insufficient knowledge and resource, the sole way to judge the truth of a statement is to consult experience. Here "experience" is used in the broad sense, not limited to personal perceptual experience only. In this situation, "truth value" and "degree of belief" are conceptually the same.

In everyday language, for a statement S, to say "S is true" is different from to say "I believe S," though their difference is not necessarily fundamental. To me, the former is like "S is not only believed by me, but also by everyone else (or that will be the case);" the latter is like "S sounds true to me, though may be not to the others." When we use the word "truth," we do imply certain *objectivity*, but it is more about "according to relatively complete and unbiased evidence" than about "as the world really is."

We can still say that in NARS "true" means "corresponds to reality," except that here reality is only revealed by the system's experience. When later we find a previous belief to be "false," it does not mean that we have had a chance to directly check the belief with reality (bypassing our experience), but that it conflicts with our updated belief based on more experience.

With such a semantics, we can still say "I strongly believe S, though it may be false," which means "I can imagine it to be rejected in the future." All of these differences cannot be used to argue that truth value cannot be the same as degree of belief.

Similarly, in NARS there is no fundamental difference among "hypothesis," "fact," "knowledge," "belief," and "guess." Instead, the difference is a matter of degree, and depends on usage convention of the words.

If we talk about such a system from an observer's point of view, then the situation is different. For example, if we have control over the experience of NARS, we may construct a situation in which the system strongly believes in a false statement. However, here "false" is from our point of view, and judged according to our knowledge about the system's future experience, which is not available to the system yet. Still, the general principle, that is, truth value is a function of experience, remains the same.

EGS challenges the traditional distinction between *ontology* and *epistemology*, that is, between *what is out there* and *what a system beliefs to be out there*. By accepting such a semantics, I do not reject the principle of naturalism — that is, the natural world exists independent of us, and it is the origin of all our knowledge [Kitchener, 1994]. What I stress here is that all *descriptions* of such an objective world in a system with insufficient knowledge and resources are intrinsically revisable. The interaction between the system and its environment is a process of assimilation and accommodation [Piaget, 1960], which usually does not maintain a one-to-one mapping between the terms/statements within the system and the objects/facts beyond the system. Actually, we cannot even talk about "objects" without assuming the cognitive capacity of some kind of system, which is what cuts reality into pieces. Since where to cut and how to cut depend on the nature and the experience of a system, there is no objective way to describe the world. What we call "ontology" is just a description of the world accepted by a community of observers, not a description of the world "as it is."

7.4.3 Validity of inference

A major motivation for the creation of EGS was to provide a justification for non-deductive inferences. As revealed by Hume's "induction problem," there is no sure way to get infallible predications about the future [Hume, 1748]. From limited past experience, we cannot get accurate descriptions of state of affairs, neither can we know how far our current belief is from such an objective description.

Based on this, Popper made the well-known conclusion that an inductive logic is impossible [Popper, 1959]. However, from the previous discussion, we can see that what is really pointed out by Hume and Popper is the impossibility of an inductive logic *with a MTS*. That is, inductive inference is invalid as far as validity is defined as "truth-preserving in all models."

If the conclusions derived in NARS are fallible, in what sense are they "better" than arbitrary guesses? This leads us to the concept of "rationality."

When infallible predictions cannot be obtained (due to insufficient knowledge and resources), beliefs based on past experience are better than arbitrary guesses, if the environment is *relatively stable* (as discussed in Section 2.1.2). To say a belief is only a summary of past experience (thus no future confirmation is guaranteed) does not make it equal to an arbitrary conclusion — it is what "adaptation" means.

Adaptation is the process in which a system changes its behaviors *as if* the future is similar to the past. It is a rational process, even though individual conclusions it produces are often wrong, and we know that the future cannot be identical to the past. For this reason, valid inference rules (deduction, induction, abduction, and so on) are the ones whose conclusions correctly (according to the semantics) summarize the evidence in the premises. They are "truth preserving" in this sense, not in the model-theoretic sense that they always generate conclusions which are immune from future revision.

7.4.4 Two types of truth

One important character of EGS is its dynamic and subjective nature. The truth value of a judgment may change from time to time in NARS, due to the arrival of new evidence. The system's inference activity also changes the truth values of judgments by combining evidence from different sections of the experience. Since truth values are based on the system's experience, they are intrinsically *subjective*. To be more precise, the system's beliefs are not objective descriptions of the world, but summaries of its own experience, so it is from the *system's point of view*. Even two systems in precisely the same environment may have different beliefs, obtained from their different individual experiences.

To say that truth values are dynamic and subjective does not mean that they are arbitrary. Different systems in the same environment can achieve a certain degree of "objectivity" by communicating to one another and thus sharing experience. However, here "objective" means "common" or "unbiased," not "observer independent." The common beliefs are still bounded by the experiences of the systems involved, though no longer by that of a single system.

The model-theoretic "truth" still has its place in NARS, though it plays a secondary role here. Whenever mathematical (or other conventional) statements are under consideration, their truth values are fixed, and are independent of the system's experience and the system's degrees of belief on them. We still do not know the truth value of Goldbach's Conjecture, though it has been confirmed in all the previous testing cases. Such a usage of the word "true" does not conflict with the fact that in the system this statement does have a truth value, calculated according to the system's experience.

From a philosophical point of view, this definition of truth is similar to Putnam's "rational acceptability" [Putnam, 1981]. In AI, a similar approach is discussed in Kowalski's paper "Logic without Model Theory," in which he defines "truth" as a relationship between sentences of the knowledge base and observational sentences [Kowalski, 1995]. However, the technical details of these approaches are quite different from NARS. For instance, Kowalski still uses predicate logic.