Experience-Grounded Semantics: A theory for intelligent systems

Pei Wang

Department of Computer and Information Sciences
Temple University, Philadelphia, PA 19122, USA
http://www.cis.temple.edu/∼pwang/
pei.wang@temple.edu

Abstract

An experience-grounded semantics is introduced for an intelligent reasoning system, which is adaptive, and works with insufficient knowledge and resources. According to this semantics, truth and meaning are defined with respect to the experience of the system — the truth value of a statement indicates the amount of available evidence, and the meaning of a term indicates its experienced relations with other terms. The major difference between experience-grounded semantics and model-theoretic semantics is that the former does not assume the sufficiency of knowledge and resources. This approach provides new ideas to the solution of some important problems in cognitive science.

Key words: model-theoretic semantics, truth and meaning, intelligent reasoning system, insufficient knowledge and resources

1 Introduction

In this paper, a new semantic theory is proposed for Artificial Intelligence (AI) reasoning systems.

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 language" or "symbolic language"), while the latter is usually evolved in history, described by some loose grammar rules that may change from time to time, and from place to place. This paper concentrates on the semantics of artificial language, though the discussion is also related to natural language.

Preprint submitted to Elsevier Science *23 August 2004*

An automatic (computerized) reasoning system usually consists of the following major components:

- a formal language for (internal) knowledge representation and (external) communication with the environment,
- a semantic theory that links the language to the environment,
- a set of inference rules that derives new knowledge from given knowledge, and answer questions according to available knowledge,
- a memory structure that stores the knowledge, questions, and intermediate results,
- a control strategy that decides what tasks to carry out and what rules to apply in each step.

Usually the first three components are referred to as consisting of a *logic*, while the last two as the *implementation* of the logic in a computer system.

In such a system, the semantic theory plays two major roles:

- It specifies how the language should be translated into other (natural or artificial) languages in communication, so that other (human or computer) systems know how to "talk" with this system. For this purpose, the semantic theory needs to specify how the *meaning* of words and sentences of the language is determined by relating them to the outside the language.
- It provides justification for the inference rules, that is, to explain why these rules, not others, are proper to be used to carry out inference on the language. For this purpose, the semantic theory needs to specify how the *truth value* of declarative sentences of the language is determined, so that the rules can be validated as preserving truth in the inference process.

The study of the semantics of formal languages has been dominated by model theory (and its variations and extensions) for decades. This paper will argue that though model-theoretic semantics is proper for many purposes, it is inappropriate for the intelligent reasoning system under discussion, where we need a fundamentally different semantic theory. In this kind of semantics, both meaning and truth are defined *with respect to the experience of the system.* Briefly speaking, an experience-grounded semantics first defines the form of experience a system can have, then defines truth value and meaning as functions of given experience.

There have been many philosophical, linguistic, and psychological discussions on truth, meaning, and semantics, as well as their relationship with experience. The aim of this paper is not to provide another argument in these debates. Instead, here we try to address the related issues from an angle which has been missing in the discussions, that is, to provide a fully formalized and computerized alternative to model-theoretic semantics. In this way, many issues in the study of semantics can be explored in a more concrete and detailed manner.

In the following, we first introduce the semantic problem we need to solve, and explain why model-theoretic semantics cannot be used there. Then, a new semantics is formally specified. Finally, this approach is compared with the traditional approach, and some of the implications of this theory are described.

2 The semantic problems in NARS

2.1 NARS overview

NARS (Non-Axiomatic Reasoning System) is an intelligent reasoning system. It is designed according to the belief that *intelligence* can be explained and reproduced as *the capacity for a system to adapt to its environment with insufficient knowledge and resources* (Wang, 1995a). In this paper, we focus on the semantics of the system. For the other aspects of the system, see the references and a demo at the author's website.

As a computerized reasoning system, NARS uses an artificial language, Narsese, to communicate with its environment. The syntax of this language is precisely specified in a formal grammar. Because NARS, in the current version, only interacts with its environment through this language, the "environment" of the system consists of a human user or another computer system. The system accepts declarative knowledge and questions (as sentences of the language) from its environment. To answer the questions according to the knowledge, it needs a memory to store them, and some inference rules to derive conclusions from given premises. These rules are formalized and built into the system when it is designed.

NARS is "intelligent" in the sense that it is adaptive, and works with insufficient knowledge and resources.

By "adaptive", we mean that NARS uses its *experience* (i.e., the history of its interaction with the environment) as the guidance of its inference activities. For each question, it looks for an answer that is most consistent with its experience (under the restriction of available resources).

By "insufficient knowledge and resources", we mean that the system has the following properties:

Finite: The system has a constant information processing capacity. As a result, it cannot be assumed that all requirements for processor time and storage space can be satisfied.

Real-time: All tasks have time requirements attached. As a result, it cannot be assumed that the system can spend as much time as it wants on a problem. Nor can it be assumed that new problems only show up when the system is idle.

Open: No constraints are put on the content of knowledge and questions that the system needs to process, as long as they are expressible in Narsese. As a result, it cannot be assumed that new knowledge will always be consistent with old knowledge. Nor can it be assumed that all required answers are deductively implied by the current knowledge.

When NARS is designed, we need a semantic theory. To communicate with NARS, we need to know how each term (or sentence) of Narsese means to the system, and how it should be understood by a human user or another computer system. To derive new knowledge from available knowledge, we need to choose inference rules whose validity can be justified. These two problems correspond to the two central issues in semantics, "meaning" and "truth", respectively. Obviously, we want the semantic theory to be consistent with the working definition of intelligence accepted in NARS, as well as to be well supported by the previous research results in cognitive science.

2.2 Model-theoretic semantics

The most natural choice of semantics in NARS is to use a model-theoretic semantics (which was indeed our initial attempt).

The basic of model-theoretic semantics 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 another language **Lm**, which is a *meta-language*, and 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 basic motivation of using formal languages is to avoid the ambiguity in natural language, so that objective and accurate artificial languages are created. Model-theoretic semantics 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 everyday 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 view point, as Tarski put it, "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 model-theoretic semantics, 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 terms and truth value of statements 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. On the other hand, what knowledge **R** has and what operations **R** performs have *no influence* on the meaning and truth value of the terms and sentences involved.

Such a treatment is desired in pure 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 ideal 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, model-theoretic semantics 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 values 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 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 model-theoretic semantics. Many variations and extensions of model-theoretic semantics 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, 1986). 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 model-theoretic semantics, 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 model-theoretic semantics is "experience independent", and it does not even need to assume the existence of a reasoning system **R** that actually uses the language.

2.3 Why NARS does not use model-theoretic semantics

Though model-theoretic semantics can be applied to NARS, it provides little help for the design and use of the system. 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", we 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 explanation should be based on the relation between the language and the world, not merely on the syntactic natures defined within the system. 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 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 the assumption of insufficient knowledge and resources, because by the very definition of *open system*, all knowledge can be challenged by future experience. 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 way. Model-theoretic semantics 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, 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 model-theoretic semantics, 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 model-theoretic semantics (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 that this semantics is accepted by the "logic-based" approach toward AI (McCarthy, 1988; Nilsson, 1991), but also it is taken to be the only possible semantics, both by its proponents and its critics. As McDermott said: "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 model-theoretic semantics is under debate, and its suitability for a natural language is controversial, few people have doubt about its suitability for a formal language. We have not seen a *formal semantics* that is not *model-theoretic*, and even such a concept may sound self-contradictory to some people.

In the following, we show that such a "non-model-theoretic formal semantics" is not only possible, but also necessary for intelligent systems.

3 An experience-grounded semantics

In this paper, we describe the semantics of NARS and a simplified version of Narsese. The other aspects of the system are only briefly introduced when necessary. For the inference rules used in NARS, see (Wang, 1994b, 2001a); for the control mechanism of the system, see (Wang, 1996c); for an overall description of the system, see (Wang, 1995a).

3.1 Inheritance and extension/intension

Narsese, the formal language used in NARS, is a *categorical language*, which is different from predicate language, because of the use of subject-predicate format in its sentences.

In its simplest form, a *term* is a string of letters in an alphabet. It corresponds to the name of a concept in NARS, and roughly to a word in a natural language.

The *inheritance relation*, "$\rightarrow$", is a relation from one term to another. It is defined by its two properties: *reflexivity* and *transitivity*. That is, for any terms x, y, and z, we always have "$x \rightarrow\ x$", and if we have "$x \rightarrow\ y$" and "$y \rightarrow\ z$", then we also have "$x \rightarrow\ z$".

An *inheritance statement* consists of two terms related by the inheritance relation. In the statement "$S \rightarrow P$", S is the *subject term* and P is the *predicate term*. "$S \rightarrow P$" means that S is a specialization of P, and P is a generalization of S. It roughly corresponds to "S is a kind of P" in English. For example, "*raven* $\rightarrow$ *bird*" corresponds to "Raven is a kind of bird".

The *idealized experience* of NARS is defined as a finite set of inheritance statements, K. It can be seen as the initial knowledge the system obtained from its interaction with the environment. The system's *idealized knowledge*, K^*, is the transitive closure of K, generated by the following algorithm:

(1) Let $K^* = K$;
(2) For each pair of propositions "$x \rightarrow y$" and "$y \rightarrow z$" in K^*, put "$x \rightarrow z$" into K^* (if it is not already there).

The last step is iterated over and over again until all possibilities have been exhausted. For a finite K, K^* is generated in finite steps.

For an arbitrary inheritance statement "$x \rightarrow y$" and given (idealized) experience K, the statement can be treated as a binary proposition (as in propositional calculus), and its *truth value* is determined by the following algorithm:

```
if $x$ and $y$ are the same term, then return $true$
else if "$x \rightarrow y$" is in $K^*$, then return $true$
else return $false$
```

Therefore, in idealized situation, there are two types of "truth" in NARS:

Analytic truth: An inheritance statement is true *by definition*, if its subject and predicate are the same term (because inheritance defined as a reflexive relation), no matter what the experience of the system is.

Synthetic truth: An inheritance statement is true *according to experience*, if its subject and predicate are different terms, and the statement is in the experience of the system, or can be derived from it by the transitivity of inheritance relation.

To specify how a particular term T is related to other terms, the *extension* and *intension* of T, relative to experience K, are defined as the sets of terms $T^E = \{x \,|\, x \rightarrow T\}$ and $T^I = \{x \,|\, T \rightarrow x\}$, respectively. That is, the extension of T includes all known specializations of T, and the intension of T includes all known generalizations of T. Extension and intension are defined in such a symmetric way that for any result about one of them, there is a corresponding result about the other. Each statement in the system's knowledge reveals part of the intension for the subject term and part of the extension for the predicate term. For example, "$S \rightarrow P$" indicates that S is in the extension of P, and P is in the intension of S.

In NARS, the *meaning* of a term T consists of its extension T^E and intension T^I, according to given experience K. Therefore, the meaning of a term is its experienced (inheritance) relations with other terms.

From the previous definitions of "inheritance", "extension", and "intension", it is not difficult to get the following result (where "$\equiv$" means "if and only if"):

$$(S \rightarrow P) \equiv (S^E \subseteq P^E) \equiv (P^I \subseteq S^I)$$

This result says that the statement "There is an inheritance relation from S to P" is equivalent to both "P inherits the extension of S" (the extension of P includes the extension of S) and "S inherits the intension of P" (the intension of S includes the intension of P). This is the reason that "$\rightarrow$" is called an *inheritance* relation. Here we use the word "inheritance" as a "two-way" relationship between two terms, and when one term get something from the other, the latter also get something else from the former.

Intuitively, such a relation indicates that one term *can be used as*, or *inherits the relations of*, the other in a certain way. If a system know "$S \rightarrow P$", then S can substitute P in sentences of the form "$P \rightarrow x$", and P can substitute S in sentences of the form "$x \rightarrow S$". The other way around, if every x that satisfies "$x \rightarrow S$" also satisfies "$x \rightarrow P$", or every x that satisfies "$P \rightarrow x$" also satisfies "$S \rightarrow x$", we have "$S \rightarrow P$". This result shows that the two major semantic notions "truth" (of statements) and "meaning" (of terms) have a close relationship in NARS.

Now we have finished our description of a simple experience-grounded semantics. For a reasoning system whose experience consists of (binary) inheritance statements, the meaning of any term and truth value of any inheritance statement can be determined according to given experience.

In the above discussion we only defined binary inheritance statements, which are either true or false. They give us a way to further define inheritance statements that are *true to a degree*. For this purpose, we first need to define *evidence*.

The previous theorem identifies an inheritance relation (from one term to another) to two subset relations (between the extensions and intensions of the two terms, respectively). Therefore, we can define "partial inheritance" by "partial subset", since the latter is already a familiar notion in set theory.

For a subset relation $S_1 \subseteq S_2$ between two sets, it is natural to define its positive evidence as elements in subset $(S_1 \cap S_2)$, and its negative evidence as elements in subset $(S_1 - S_2)$.

Since inheritance is about both extension and intension, we define (positive and negative) evidence for an inheritance statement "$S \rightarrow P$" as the following:

- A piece of *positive* evidence is a term M such that $M \in (S^E \cap P^E)$ or $M \in (P^I \cap S^I)$.
- A piece of *negative* evidence is a term M such that $M \in (S^E - P^E)$ or $M \in (P^I - S^I)$.

The intuition behind the above definition is: since "$S \rightarrow P$" states that "S inherits the intension of P, and P inherits the extension of S", then M is a piece of positive evidence if as far as M is concerned the inheritance is true, and M is a piece of negative evidence if as far as M is concerned the inheritance is false.

To measure the amount of evidence, here we simply use the size of the corresponding set. Therefore, we have

$$w^+ = |S^E \cap P^E| + |P^I \cap S^I|,$$

$$w^- = |S^E - P^E| + |P^I - S^I|,$$

$$w = w^+ + w^- = |S^E| + |P^I|.$$

Here w^+ and w^- are the amounts of positive and negative evidence, respectively. Their sum, w, is the amount of all available evidence for "$S \rightarrow P$".

Though in principle all the information that we want to put into a truth value is representable by any two of the above three values, it is not always natural or convenient for the purpose of NARS. Instead of using "absolute measurements" as truth value, we often prefer "relative measurements", such as real numbers in the interval [0, 1].

A natural relative measurement is the *frequency* (or *proportion*) of positive evidence among all evidence, $f = w^+/w$. Because w is the number of times that the proposed inheritance statement is checked, and w^+ is the number of times that the statement is confirmed, f indicates the "success frequency" of the inheritances (of extension and intension) between the two terms, according to the experience of the system.

To represent a truth value by a frequency value alone is not enough for NARS. As an open system, it also needs a way to measure the stability of the frequency measurement with respect to future evidence. For this purpose, we compare the current amount of evidence, w, to a constant amount of future evidence, k, and define the *confidence* of a statement as $c = w/(w + k)$.

Intuitively, confidence is the ratio of the amount of the "all current evidence" to the amount of the "all evidence in the near future". It indicates how much the system knows about the inheritance relation. Since k is a constant, the more the system knows about the inheritance relation (i.e., the bigger w is), the more confident the system is about the frequency, since any effect of the evidence arriving in the near future will be relatively smaller. For our current purposes, k can be any positive number, and in the current implementation, $k = 1$ is used as a default. For a detailed discussion on confidence, see (Wang, 2001b).

From the above definitions, it is easy to see that the two measurements, f and c, are *independent* of each other, in the sense that from the value of one, the value of the other cannot be determined, or even bounded. Therefore, in NARS the truth value of a statement is represented by a pair of real numbers in [0, 1]. A statement plus its truth value is called a *judgment*, and it has the form " $S \rightarrow P <f, c>$".

Because NARS is designed under the assumption of insufficient knowledge and resources, all the judgments within the system are supported by finite evidence — that is, w is positive and finite, so $0 < c < 1$. Beyond the normal truth values, there are two limiting cases useful for the interpretation of truth values and the justification of inference rules:

Null evidence: This is represented by $c = 0$ (or $w = 0$). It means that the system knows nothing at all about the inheritance relation (so f is undefined).

Full evidence: This is represented by $c = 1$ (or $w = \infty$). It means that the system already has complete information about the statement (so no future modification of the truth value is possible).

Especially, "$S \rightarrow P < 1, 1 >$" is exactly the binary inheritance statement "$S \rightarrow P$" we introduced previously. In this special case, the effects of negative evidence and future evidence can be ignored, and the inheritance relation is "complete".

Now let us see the whole picture described so far: NARS uses a formal language, Narsese, for knowledge representation and communication, and the sentences of the language are inheritance judgments of the form "$S \rightarrow P < f, c >$" (more complicated statements will be introduced later). Obviously, now whether a term x is in the extension or the intension of T is also a matter of degree, measured by a truth value as defined above.

Therefore, Narsese has been given an experience-grounded semantics. A subset of Narsese, in which all judgments have truth value $< 1, 1 >$, is used to represent the idealized experience K. According to a given K, the truth value $<f, c>$ of any given statement "$S \rightarrow P$" can be evaluated (by comparing the extensions and intensions of S and P), and the meaning of any given term T can be determined (by evaluating its extensional statements "$x \rightarrow T$" and intensional statements "$T \rightarrow x$" for every term x in K).

3.3 Ideal and actual experience

As mentioned at the beginning of the paper, the above defined experience-grounded semantics plays two major roles in the design and use of NARS: language understanding and rule justification.

If "*raven* $\rightarrow$ *black-thing* $< 9/10, 10/11 >$" is a piece of knowledge within NARS, then it indicates that the system believes the inheritance statement "*raven* $\rightarrow$ *black-thing*" to the extent that *as if* the statement has been tested 10 times in idealized situations, and in 9 of them the relation is confirmed (with the system parameter k set to 1). Also, this knowledge contributes to the meaning of terms "*raven*" (by indicating that "*black-thing*" is in its intension to the extent measured by $<9/10, 10/11>$) and "*black-thing*" (by indicating that "*raven*" is in its extension to the extent measured by $<9/10, 10/11>$). This does not imply, of course, that the system actually got the truth value by carrying out the testings — such "ideal evidence" cannot be obtained practically. Indeed, the system may have checked the relation more than 10 times, or the conclusion may have been derived from other knowledge or even directly provided by the environment. But no matter how the truth value "$<9/10, 10/11>$" is generated in practice (there are infinitely many ways it could arise), it can always be *understood* in a unique way, as stated above.

Since truth value is defined with respect to available evidence, in NARS the validity of inference rules can still be defined as "truth preserving", but now in the sense that the truth value of the conclusion is evaluated according to the evidence provided by the premises. For example, in NARS the induction rule derives "*bird* $\rightarrow$ *black-thing* $<1,\ 1/2>$" from "*raven* $\rightarrow$ *bird* $<1,\ 1>$" and "*raven* $\rightarrow$ *black-thing* $<1,\ 1>$", because the premises indicate that "*raven*" provides a piece of positive evidence for the conclusion, and this is indeed the information embedded in the truth value of the conclusion. In this way, the semantics (the definitions of evidence and truth value) provides constraints for the inference rules, where truth-value functions calculate the truth value of the conclusion from the truth values of the premises. For detailed discussion about the inference rules, see (Wang, 1994b, 2001a).

In NARS, ideal experience is used to *define* meaning and truth value and to *design* inference rules, while actual experience is used to *determine* truth value and meaning. Such a definition is desired, because, as Krantz put it, "numerical statements are meaningful insofar as they can be translated, using the mapping conventions, into statements about the original qualitative structure" (Krantz, 1991). In other words, the "ideal experience" is used here as an "ideal meter-stick" to measure the degree of truth. Like all measurements, though its unit is defined in an idealized situation, it is not used only in idealized situations. In (Putnam, 1981), Putnam treated truth as only defined under "epistemically ideal conditions", but can be used in non-ideal situations. He compared this treatment with the notion of a "frictionless plane", which is an important concept in physics, though cannot be practically obtained. In NARS, "truth" is a statement with truth value $<1,\ 1>$. By definition, it cannot be obtained in an open system like NARS, though it can be infinitely approached by actual knowledge, as well as used in the semantic foundation of the system.

We cannot directly use the actual experience, which is a sequence of inheritance judgments, to define meaning and truth value, because that would cause circular definition — the judgments in experience have truth values themselves. What we have done is to first build the semantics for a special subset of the language where truth value is binary, then use it to build the semantics of the entire language. In this way, each actual input judgment is seen as corresponding to a set of idealized judgments. To define truth values and meanings in terms of experience does not give the judgments in the system's actual experience any privilege. Just like any others, their truth value can also be changed by other sections of the system's experience.

Another factor that makes actual experience different from ideal experience is the insufficiency of resources. Due to the lack of memory space, not all of the system's experience will be remembered; due to the lack of processing time, some remembered judgments will be ignored. Consequently, the truth value of a sentence or the explicit meaning of a term is usually based on *partial experience*, or a *section* of the system's actual experience.

Working under time pressure, NARS does not attempt to judge the truth value of a statement according to all available (relevant) knowledge. Instead, the system allocates a certain amount of time for each task (such as a question to be answered, or a piece of new knowledge to be digested) according to the current resource request/supply situation (such as how many other tasks are under processing), and reports the best (e.g., most confident) answer it finds for each question with the given knowledge and resources.

For the same reason, when a term is used by the system to answer a question, only some, but not all, of its relations with other terms are taken into account. The system maintains a priority ranking among the relations, according to how important and relevant they are. The system gives relations with high priority more chance to be used, and adjusts the priority distribution according to the feedback from processing results (Wang, 1995a, 1996c).

Therefore, for an actual judgment in NARS, its truth value is understood in terms of idealized experience, but the value itself either directly comes from the environment (in that case, it is assigned by the user or another system according to the semantics of NARS) or comes from certain inference rule (in that case, it is calculated according to the semantics of NARS from other available judgments). For an actual term in NARS, its meaning is indicated by its available relations with other terms. In this way, both truth and meaning are eventually determined by the system's actual experience, though usually in a complicated manner.

3.4 Basic inference rules

NARS uses *syllogistic* inference rules. A typical syllogistic rule takes two judgments sharing a common term as premises, and derives a conclusion, which is a judgment between the two unshared terms. For inference among inheritance judgments, there are three possible combinations if the two premises share exactly one term:

$$\{M \rightarrow P <f_1, c_1>,\ S \rightarrow M <f_2, c_2>\} \vdash S \rightarrow P <F_{ded}>$$
$$\{M \rightarrow P <f_1, c_1>,\ M \rightarrow S <f_2, c_2>\} \vdash S \rightarrow P <F_{ind}>$$
$$\{P \rightarrow M <f_1, c_1>,\ S \rightarrow M <f_2, c_2>\} \vdash S \rightarrow P <F_{abd}>$$

The three rules above correspond to *deduction*, *induction*, and *abduction*, respectively, as indicated by the names of the truth-value functions. In each of these rules, the two premises come with truth values $<f_1, c_1>$ and $<f_2, c_2>$, and the truth value of the conclusion, $<f, c>$, is a function of them — according to the experience-grounded semantics, the truth value of the conclusion is evaluated with respect to the evidence provided by the premises.

These truth-value functions are designed in the following procedure:

(1) Treat all relevant variables as binary variables taking 0 or 1 values, and determine what values the conclusion should have for each combination of premises, according to the semantics.
(2) Represent the variables of conclusion as Boolean functions of those of the premises, satisfying the above conditions.
(3) Extend the Boolean operators into real number functions defined on [0, 1] in the following way:

$$\begin{aligned} not(x) &= 1 - x \\ and(x_1, ..., x_n) &= x_1 * ... * x_n \\ or(x_1, ..., x_n) &= 1 - (1 - x_1) * ... * (1 - x_n) \end{aligned}$$

(4) Use the extended operators, plus the relationship between truth value and amount of evidence, to rewrite the functions as among truth values (if necessary).

For the above rules, the resulting functions are:

$$\begin{aligned} &F_{ded} : f = f_1 f_2 && c = f_1 f_2 c_1 c_2 \\ &F_{ind} : f = f_1 && c = f_2 c_1 c_2 / (f_2 c_1 c_2 + k) \\ &F_{abd} : f = f_2 && c = f_1 c_1 c_2 / (f_1 c_1 c_2 + k) \end{aligned}$$

When two premises contain the same statement, but comes from different sections of the experience, the revision rule is applied to merge the two into a summarized conclusion:

$$\{S \rightarrow P <f_1, c_1>,\ S \rightarrow P <f_2, c_2>\} \vdash S \rightarrow P <F_{rev}>$$

$$F_{rev} : f = \frac{f_1 c_1/(1-c_1) + f_2 c_2/(1-c_2)}{c_1/(1-c_1) + c_2/(1-c_2)} \quad c = \frac{c_1/(1-c_1) + c_2/(1-c_2)}{c_1/(1-c_1) + c_2/(1-c_2) + 1}$$

The above function is derived from the additivity of the amount of evidence and the relation between truth value and amount of evidence.

The revision rule can be used to merge less confident conclusions, so as to get more confident conclusions. In this way, patterns repeatedly appear in the experience can be recognized and learned.

3.5 Other types of statement in Narsese

The inheritance statement "$S \rightarrow P$" is the basic form of statement in Narsese, but it is not the only form. In addition to it, the language contains other types of statements, which are all built upon this basic form.

Derived inheritance relations: Beside the inheritance relation defined previously, NARS also includes several of its variations. For example,

- The *similarity* relation "$\leftrightarrow$" is symmetric inheritance. For example, "$raven \leftrightarrow crow$" means "Raven is similar to crow";
- The *instance* relation "$\circ\!\!\rightarrow$" is an inheritance relation where the subject term is treated as an atomic instance of the predicate term. For example, "$Tweety \circ\!\!\rightarrow bird$" means "Tweety is a bird";
- The *property* relation "$\rightarrow\!\!\circ$" is an inheritance relation where the predicate term is treated as a primitive property of the subject term. For example, "$raven \rightarrow\!\!\circ black$" means "Ravens are black".

Compound terms: In inheritance statements, the (subject and predicate) terms not only can be simple terms (as in the above examples), but also can be compound terms formed by other terms with a logical operator. For example, if A and B are terms, then

- their *extensional intersection* $(A \cap B)$ is a compound term, initially defined by $(A \cap B)^E = (A^E \cap B^E)$ and $(A \cap B)^I = (A^I \cup B^I)$.
- their *intensional intersection* $(A \cup B)$ is a compound term, initially defined by $(A \cup B)^I = (A^I \cap B^I)$ and $(A \cup B)^E = (A^E \cup B^E)$;

Initially, the meaning of a compound is determined according to its logical relation with its components (its "definition"), but as soon as the system begins to get (input or derived) judgments on the compound, they also contribute to its meaning, and such contributions cannot always be reduced to its components. Therefore the *principle of compositionality* in semantics is only partially true in NARS.

Ordinary relation: In NARS, only the inheritance relation and its variations are defined as logic constants that are directly recognized by the inference rules. All other relations are converted into inheritance relations with compound terms. For example, an arbitrary relation R among three terms A, B, and C is usually written as $R(A, B, C)$, which can be equivalently rewritten as one of the following inheritance statements (i.e., they have the same meaning and truth value):

- $(\times\ A\ B\ C) \rightarrow R$, where the subject term is a compound $(\times\ A\ B\ C)$. This statement says "The relation among A, B, C (in that order) is an instance of the relation R."
- $A \rightarrow (\perp\ R\ \diamond\ B\ C)$, where the predicate term is a compound $(\perp\ R\ \diamond\ B\ C)$ with a "wildcard", $\diamond$. This statement says "A is such an x that satisfies $R(x, B, C)$."
- $B \rightarrow (\perp\ R\ A\ \diamond\ C)$. Similarly, "$B$ is such an x that satisfies $R(A, x, C)$."

- $C \rightarrow (\perp\ R\ A\ B\ \diamond)$. Again, "$C$ is such an x that satisfies $R(A, B, x)$."

Higher-order term: In NARS, a statement can be used as a term, and called a "higher-order" term. For example, "Bird is a kind of animal" is represented by statement "*bird* → *animal*", and "People know that bird is a kind of animal" is represented by statement "$(bird \rightarrow animal) \circ\!\!\rightarrow (\perp\ know\ people\ \diamond)$", where the subject term is a statement. Compound higher-order terms are also defined: if A and B are higher-order terms, so do their negations ($\neg A$ and $\neg B$), disjunction ($A \vee B$), and conjunction ($A \wedge B$).

Higher-order relation: Higher-order relations are the relations whose subject term and predicate term are both higher-order terms. In NARS, there are two of them defined as logic constants:

- *implication*, "⇒", which intuitively corresponds to "if-then", and is defined as isomorphic to *inheritance*, "→";
- *equivalence*, "⇔", which intuitively corresponds to "if-and-only-if", and is defined as isomorphic to *similarity*, "↔".

For each type of terms/statements, its meaning/truth-value is defined similarly to how we define meaning/truth-value for term/statement in inheritance relation. There are inference rules taking these statements as premises or conclusions. Detailed discussion about them is beyond the scope of this paper. They are mentioned here merely to show that the proposed semantics does not only support a simple formal language.

With the above term/statement types, the expressive and inferential power of Narsese is greatly enriched. There is no one-to-one mapping between sentences in this language and those in first-order predicate calculus, though approximate mapping is possible for many sentences. While first-order predicate calculus may still be better for representing mathematical knowledge, this new language will be better for representing empirical knowledge, partially because of its experience-grounded semantics.

4 Comparison and discussion

Experience-grounded semantics is different from model-theoretic semantics in several important aspects, and their comparison is related to many issues in cognitive science. In this section we only discuss some of them.

4.1 Experience vs. model

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.
- A model must be consistent, whereas judgments in experience may conflict with one another.
- A model of language **L** is described in another language **Lm**, whereas experience can be 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, an experience must be the experience of a system.

These two types of descriptions serve different purposes. In general, we can distinguish two types of reasoning systems:

Axiomatic system: All inference processes in the system start from a constant set of (consistent) axioms. Whether a statement is a theorem depends on whether it can be proved according to the axioms, and how many steps the proof needs does not matter.

Non-axiomatic system: New knowledge is added into the system from time to time. The system has to answer questions according to available knowledge, which may be incomplete and inconsistent. The inference process is bounded by the available time-space of the system.

Clearly, an axiomatic system assumes the sufficiency of knowledge and resources with respect to the questions to be answered, and 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. On the contrary, a non-axiomatic system assumes the insufficiency of knowledge and resources with respect to the questions to be answered, and always attempts to answer a question with available knowledge and resources, which means that the system may revise its beliefs from time to time.

Generally speaking, human mind works with insufficient knowledge and resources (Medin and Ross, 1992). Therefore, human inference process is more similar to that of a non-axiomatic system, than to an axiomatic one. However, for certain relatively mature and stable knowledge, it is more efficient to treat them as an axiomatic (sub)system (within the whole non-axiomatic system). 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 postulations and conventions. After we figure out the implications of these postulations and conventions, we can apply such a theory into many situations, because as far as the postulations and conventions can be "instantiated" by substituting the abstract concepts with the concrete ones, all the ready-made implications follows. This is the picture provided by model-theoretic semantics.

On the other hand, if the knowledge embedded in a reasoning system is not mathematical, but empirical, then what we have is fundamentally a non-axiomatic system, in which 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 experience-grounded semantics 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, to turn the problem of word meaning into the problem of concept meaning. The meaning of concepts is not simpler than the meaning of words 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, 1995). These issues are hard to handle in model-theoretic semantics. In NARS, since a term is the name of a concept, the meaning/truth defined for the language and the meaning/truth defined for the concepts system are isomorphic to each other, so the current discussion applies to both.

Though overall NARS uses experience-grounded semantics, there are still places where model-theoretic semantics is used. One example is the variables in the inference rule. As mentioned previously, the induction rule of NARS derives "$S \to P < f, c >$" from "$M \to P < f_1, c_1 >$" and "$M \to S < f_2, c_2 >$". Written in this way, the variables S, M, and P have no experience-related meaning until they are instantiated by constant terms "*bird*", "*raven*", and "*black-thing*", respectively, and then the meaning of the variables is determined by the meaning of the constants. Similarly, when mathematical knowledge is provided to NARS, it will be used with model-theoretic semantics.

4.2 Language and uncertainty

As mentioned previously, Narsese is a categorical language, in which each statement consists of a subject term and a predicate term, related together by an inheritance relation (or its variations). This type of language is exemplified by Aristotle's syllogism (Aristotle, 1989; Bocheński, 1970; Englebretsen, 1981).

In NARS, inheritance relations and its variations (similarity, instance, property, implication, and equivalence) are logical constants, whose meaning is fixed, and independent of the system's experience, while the meaning of the other relations (ordinary relations, such as "part of", "between", "younger than", "know", "believe") is defined by their extension and intension, and therefore is experience-dependent.

The meaning of inheritance relation is closely related to many well-known relations — for instance, "is-a" (in semantic networks) (Brachman, 1983), "belongs to" (in Aristotle's syllogisms), "subset" (in set theory), "inheritance assertion" (Touretzky, 1986), as well as many relations studied in AI, computer science, psychology, and philosophy, such as "type–token", "category–instance", "class–object", "general–specific", and "superordinate–subordinate". What makes inheritance (as defined in NARS) different from the others is: it is a relation between two *terms*, and the relation is completely defined by the two properties: *reflexivity* and *transitivity.*

A comprehensive comparison between NARS and first-order predicate calculus is beyond the scope of this paper. For the discussion of semantics, here we only explain one reason for NARS to abandon predicate calculus, that is, predicate calculus does not handle the concept "evidence" properly.

For example, "Ravens are black" is represented in first-order predicate calculus as a universally quantified proposition: $(\forall x)(Raven(x) \rightarrow Black(x))$. A natural idea, "Nicod's Criterion", is to take black ravens as its positive evidence, non-black ravens as its negative evidence, and non-ravens (black or not) as irrelevant. However, this approach leads to Hempel's "Confirmation Paradox" (Hempel, 1943). According to Hempel, since $(\forall x)(\neg Black(x) \rightarrow \neg Raven(x))$ is equivalent to $(\forall x)(Raven(x) \rightarrow Black(x))$. Non-black non-ravens (such as a green shirt) are positive evidence of the former, and therefore are also positive evidence of the latter. This conflicts with Nicod's Criterion.

Though almost all previous discussions (e.g. Swinburne, 1973) treat this problem as a *paradox of logic*, it is actually a *paradox of first-order predicate logic.* It appears because in first-order predicate logic all general statements contains universally quantified variables which can be instantiated by any constant in the domain, so *nothing is irrelevant*, and a general statement is always talking about everything in the domain. Furthermore, only negative evidence contribute to the truth value of such a proposition.

It is no longer the case in a categorical language like Narsese. Here for "Ravens are black", black ravens are positive evidence, non-black ravens are negative evidence, and non-ravens are not directly relevant. On the other hand, for "non-black things are not ravens", non-black non-ravens (including green shirts) are positive evidence, non-black ravens are negative evidence, and ravens are not directly relevant. So in NARS the two statements have the same negative evidence but different positive evidence. In first-order predicate logic, truth value only depends on the existence of negative evidence, so these two statements are equivalent, and Hempel's paradox follows. On the contrary, since in NARS truth value is determined both by positive and negative evidence, these two statements are no longer equivalent, and may have different truth values. Consequently, the paradox does not appear in NARS.

A related issue is why NARS does not directly use probability theory to represent and calculate a truth value. This issue has been discussed in previous publications (Wang, 1993, 1996a, 2001b), so here we only briefly address it. From our previous definition of truth value, it is easy to recognize its relationship with probability theory. Intuitively speaking, the frequency measurement is similar to probability, and the confidence measurement is related to the size of sample space, so that each judgment corresponds to a probability distribution function. However, truth values of different judgments cannot been treated as belonging to the same probability distribution, because each of them has its own evidence space (defined by the extension of its subject and the intension of its predicate), and its truth value is evaluated according to different body of evidence. If we use the terminology of probability theory, then the truth value functions in NARS correspond to cross-distribution calculations, which are not specified in standard probability theory.

There are other publications comparing NARS with other uncertainty representation and processing approaches, such as fuzzy logic (Wang, 1996b), Dempster-Shafer theory (Wang, 1994a), and non-monotonic logics (Wang, 1995b). In general, NARS is different from them, because none of the existing approaches fully satisfies the assumption of insufficient knowledge and resources, so none of them can be used with an experience-grounded semantics.

4.3 On meaning

The definition of meaning in the experience-grounded semantics of NARS 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 nothing 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. 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 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.

By saying so, we do not mean that 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 experience-grounded semantics, all the symbols that the system has 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 sensory–motor 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 sensory–motor 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 means "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".

The definition of meaning in NARS is similar to 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 experience-grounded semantics and those theories are:

- In NARS, the relations among terms are not definitional or linguistic, but *experienced* relations that happen in 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 relations, 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.

Traditionally, *extension* and *intension* refer to two aspects of the meaning of a term: roughly speaking, its *instances* and its *properties*. A term's extension is usually defined as a set of *objects* in a "physical world" that are denoted by the given term; the term's intension is usually defined as a *concept* in a "Platonic world" which denotes or describes the given term (Bocheński, 1970; Inhelder and Piaget, 1969; Kitchener, 1994). 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. However, in the current theory, they are defined using (the two sides of) an inheritance relation between two terms, which is *within* the language, yet even so, the definition still keeps the intuitive quality that "extension" refers to instances, and "intension" refers to properties.

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 experience-grounded semantics, the meaning of a term is not *reduced into 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 inheritance relation and its variations are logical constants in the language. Their meaning is innate to the system, because they are directly recognized by the inference rules and control mechanism. Even when all the terms in an input statement are novel, the inheritance relation is known. Therefore, NARS is not "getting meaning out of meaningless".

As mentioned previously, due to insufficient resources, the system cannot consult all known judgments associated with a term each time the term is used. Instead, in NARS a *priority* distribution is maintained among these judgments, which determines the chance for a certain judgment to be taken into consideration at a certain time. The distribution is adjusted by the system according to the feedback of each inference step (to make the more useful knowledge more accessible), as well as according to the current context (to make the more relevant knowledge more accessible). Consequently, the meaning of a term become 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 judgments are ranked in terms of their priority. When the system gets new knowledge, or turns to another question, 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 gradually know how the term is used by the other. In the long run, 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, therefore have shared experience.

This conclusion to an extent agree 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). In this way, the semantics of NARS also implies a new theory of categorization, which is discussed in (Wang, 2002).

4.4 On truth

As defined previously, in NARS "truth" corresponds to statements with truth value $< 1, 1 >$, and it can only be approached, but not reached, by actual knowledge 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 experience-grounded semantics of NARS has the following implications:

(1) Truth is a matter of degree, and is determined by the extent to which the two terms in 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, and the other for the relative amount of all available evidence.
(3) A truth value is assigned to a statement according to the *past* experience of the system. It does not indicates whether the statement will be consistent with *future* experience, though an adaptive system behaves according to it.
(4) The truth value of a statement may change within a system according to experience and context.
(5) Truth is subjective, but not arbitrary.

Model-theoretic semantics provides a "correspondence" theory of truth, where the truth value of a statement is determined by whether it agrees with the world, as described in a meta-language. According to model-theoretic semantics, "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 who 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. 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 "from every people's point of view" than about "as the world really is". We can still say that in NARS "true" means "agree with 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 image 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.

If we talk about such a system from an observer's point of view, then the situation may be 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.

By accepting such a semantics, we do not reject the principle of naturalism — that is, the natural world, with objects and relations among them, exists independent of us, and it is the origin of all our knowledge (Kitchener, 1994). What we 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.

Such a semantics provides a justification for non-deductive inferences. As revealed by Hume's "induction problem", our predications about future experience cannot be infallible (Hume, 1748). From limited past experience, we cannot get accurate descriptions of "state of affairs", neither can we know how far our current knowledge 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 model-theoretic semantics*.

If the answers provided by NARS are fallible, in what sense these answers are "better" than arbitrary guesses? This leads us to the concept of "rationality". When infallible predictions cannot be obtained (due to insufficient knowledge and resources), answers based on past experience are better than arbitrary guesses, if the environment is *relatively stable*. To say an answer is only a summary of past experience (thus no future confirmation 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. 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.

An important nature of the definition of truth value in NARS is that the "extensional factor" and the "intensional factor" are merged. Although it is possible to develop extensional logics or intensional logics separately (Wang, 1994b), we also need systems that can mix them together, because the coordination of the extension and intention 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, extensional evidence and intensional evidence are often compared and combined. We may 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*. Having defined extension and intension as each other's "dual" in the previous section, we now have good reason to treat them uniformly in our development of NARS.

One important character of experience-grounded semantics is its dynamic and subjective nature. The truth value of a judgment changes 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 knowledge is not an objective description of the world, but a summary 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 knowledge, 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 knowledge 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. Similarly, we can let NARS say "I don't know whether S is true" whenever in the system statement S has a low confidence value. Such a usage of the word "true" does not conflict with the fact that in the system S 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, 1994). However, the technical details of these two approaches are quite different. For instance, Kowalski still uses first-order predicate logic.

5 Conclusion

In this paper, we introduces a new semantic theory, in which meaning and truth are defined according to the experience of a system. This approach is fundamentally different from model-theoretic semantics. The former assumes insufficient knowledge and resources, while the latter assumes sufficient knowledge and resources. The former is mostly for empirical science and everyday thinking, while the latter is mostly for mathematics.

Since these two types of semantics are based on different assumptions and serve different purposes, they are not really competitors. The current problem is that many people take model-theoretic semantics as the only possible semantics, and apply it to situations where its assumptions cannot be satisfied. Therefore, in a practical sense, this new semantics is competing with model-theoretic semantics in AI and cognitive science. In this kind of fields, an experience-grounded semantics may be more fruitful than a model-theoretic semantics.

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. We do not present it 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, we do believe that in general this model is closer to a descriptive model than model-theoretic semantics is. Though it is not the major goal of the current research, it will be interesting to explore the implications of this theory in philosophy, linguistics, and psychology.

Human beings often (if not always) judge the truth value of a statement according to experience and determine the meaning of a word according to its relations with other words. This is not a novel idea to philosophers, psychologists, and linguists. However, few people tried to apply it to an *artificial* language defined by a formal grammar. People often implicitly assume that the semantics of a formal language has to be model-theoretic. Actually, a language can be “formal” in two different senses. In a weak sense, it means that the language is artificial, and formed according to a formal grammar; in a strong sense, it means that the language is also used with a model-theoretic semantics. Narsese is “formal” in the weak sense only. In this paper, we show that it is possible to give the language a non-model-theoretic formal semantics.

The semantics of NARS is not the only possible experience-grounded semantics. For systems designed for different purposes, other instances of this kind of semantics can be developed. Though they may have different forms, they should share the same theoretical foundation, that is, semantic notions of a language, such as meaning and truth, are defined as functions of the experience of a system using the language.

The NARS project is based on the belief that “intelligence” should be treated as the capacity of adaptation with insufficient knowledge and resources. In this sense, the experience-grounded semantics introduced here is the semantics for intelligent systems. From the above discussion, we can see that such a semantics addresses many important problems in a consistent manner, and suggests a new direction for AI and cognitive science.

Acknowledgment

Thanks to Ben Goertzel and John Millaway for their helpful comments.

References

Aristotle, 1989. Prior Analytics. Hackett Publishing Company, Indianapolis, Indiana, translated by R. Smith.

Barsalou, L., 1999. Perceptual symbol systems. Behavioral and Brain Sciences 22, 577–609.

Barwise, J., Perry, J., 1983. Situations and Attitudes. MIT Press, Cambridge, Massachusetts.

Birnbaum, L., 1991. Rigor mortis: a response to Nilsson's "Logic and artificial intelligence". Artificial Intelligence 47, 57–77.

Bocheński, I., 1970. A History of Formal Logic. Chelsea Publishing Company, New York, translated and edited by I. Thomas.

Brachman, R., 1983. What is-a is and isn't: an analysis of taxonomic links in semantic networks. IEEE Computer 16, 30–36.

Brooks, R., 1991. Intelligence without representation. Artificial Intelligence 47, 139–159.

Carnap, R., 1950. Logical Foundations of Probability. The University of Chicago Press, Chicago.

Dummett, M., 1978. Truth and Other Enigmas. Harvard University Press, Cambridge, Massachusetts.

Ellis, J., 1993. Language, Thought, and Logic. Northwestern University Press, Evanston, Illinois.

Englebretsen, G., 1981. Three Logicians. Van Gorcum, Assen, The Netherlands.

Field, H., 2001. Truth and the Absence of Fact. Oxford University Press, New York.

Fodor, J., 1987. Psychosemantics. MIT Press, Cambridge, Massachusetts.

Halpern, J., 1990. An analysis of first-order logics of probability. Artificial Intelligence 46, 311–350.

Harman, G., 1982. Conceptual role semantics. Notre Dame Journal of Formal Logic 28, 252–256.

Harnad, S., 1990. The symbol grounding problem. Physica D 42, 335–346.

Hempel, C., 1943. A purely syntactical definition of confirmation. Journal of Symbolic Logic 8, 122–143.

Hofstadter, D., FARG, 1995. Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought. Basic Books, New York.

Hume, D., 1748. An Enquiry Concerning Human Understanding. London.

Inhelder, B., Piaget, J., 1969. The Early Growth of Logic in the Child. W. W. Norton & Company, Inc., New York, translated by E. Lunzer and D. Papert.

Kitchener, R., 1994. Semantic naturalism: The problem of meaning and naturalistic psychology. In: Overton, W., Palermo, D. (Eds.), The Nature and Ontogenesis of Meaning. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Kowalski, R., 1994. Logic without model theory. In: Gabbay, D. M. (Ed.), What is a Logical System? Oxford University Press, pp. 35–71.

Krantz, D., 1991. From indices to mappings: The representational approach to measurement. In: Brown, D., Smith, J. (Eds.), Frontiers of Mathematical Psychology: Essays in Honor of Clyde Coombs. Recent Research in Psychology. Springer-Verlag, Berlin, Germany, Ch. 1.

Kyburg, H., 1992. Semantics for probabilistic inference. In: Proceedings of the Eighth Conference on Uncertainty in Artificial Intelligence. pp. 142–148.

Lakoff, G., 1987. Women, Fire, and Dangerous Things: What Categories Reveal about the Mind. University of Chicago Press, Chicago.

Lakoff, G., 1988. Cognitive semantics. In: Eco, U., Santambrogio, M., P., V. (Eds.), Meaning and Mental Representation. Indiana University Press, Bloomington, Indiana.

Lynch, M., 1998. Truth in Context. MIT Press, Cambridge, Massachusetts.

McCarthy, J., 1988. Mathematical logic in artificial intelligence. Dædalus 117 (1), 297–311.

McDermott, D., 1987. A critique of pure reason. Computational Intelligence 3, 151–160.

Medin, D., Ross, B., 1992. Cognitive Psychology. Harcourt Brace Jovanovich, Fort Worth.

Nilsson, N., 1991. Logic and artificial intelligence. Artificial Intelligence 47, 31–56.

Palmer, F., 1981. Semantics, 2nd Edition. Cambridge University Press, New York.

Piaget, J., 1960. The Psychology of Intelligence. Littlefield, Adams & Co., Paterson, New Jersey.

Popper, K., 1959. The Logic of Scientific Discovery. Basic Books, New York.

Putnam, H., 1981. Reason, Truth and History. Cambridge University Press, Cambridge.

Quillian, M. R., 1968. Semantic memory. In: Minsky, M. (Ed.), Semantic Information Processing. MIT Press, Cambridge, Massachusetts.

Russell, B., 1901. Recent work on the principles of mathematics. International Monthly 4, 83–101.

Segal, G., 2000. A Slim Book about Narrow Content. MIT Press, Cambridge, Massachusetts.

Smolensky, P., 1988. On the proper treatment of connectionism. Behavioral and Brain Sciences 11, 1–74.

Swinburne, R., 1973. An Introduction to Confirmation Theory. Methuen, London.

Tarski, A., 1944. The semantic conception of truth. Philosophy and Phenomenological Research 4, 341–375.

Touretzky, D., 1986. The Mathematics of Inheritance Systems. Pitman Publishing, London.

van Gelder, T., 1997. Dynamics and cognition. In: Haugeland, J. (Ed.), Mind Design II. MIT Press, Cambridge, Massachusetts, pp. 421–450.

Wang, P., 1993. Belief revision in probability theory. In: Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers, San Mateo, California, pp. 519–526.

Wang, P., 1994a. A defect in Dempster-Shafer theory. In: Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers, San Mateo, California, pp. 560–566.

Wang, P., 1994b. From inheritance relation to nonaxiomatic logic. International Journal of Approximate Reasoning 11 (4), 281–319.

Wang, P., 1995a. Non-axiomatic reasoning system: Exploring the essence of intelligence. Ph.D. thesis, Indiana University.

Wang, P., 1995b. Reference classes and multiple inheritances. International Journal of Uncertainty, Fuzziness and and Knowledge-based Systems 3 (1), 79–91.

Wang, P., 1996a. Heuristics and normative models of judgment under uncertainty. International Journal of Approximate Reasoning 14 (4), 221–235.

Wang, P., 1996b. The interpretation of fuzziness. IEEE Transactions on Systems, Man, and Cybernetics 26 (4), 321–326.

Wang, P., 1996c. Problem-solving under insufficient resources. In: Working Notes of the AAAI Fall Symposium on Flexible Computation. Cambridge, Massachusetts, pp. 148–155.

Wang, P., 2001a. Abduction in non-axiomatic logic. In: Working Notes of the IJCAI workshop on Abductive Reasoning. Seattle, Washington, pp. 56–63.

Wang, P., 2001b. Confidence as higher-order uncertainty. In: Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications. Ithaca, New York, pp. 352–361.

Wang, P., 2002. The logic of categorization. In: Proceedings of the 15th International FLAIRS Conference. Pensacola, Florida, pp. 181–185.

Whorf, B., 1956. Language, Thought, and Reality. MIT Press, Cambridge, Massachusetts.

Wittgenstein, L., 1999. Philosophical Investigations. Prentice Hall, Upper Saddle River, New Jersey, translated by G. Anscombe.

Wright, C., 1992. Truth and Objectivity. Harvard University Press, Cambridge, Massachusetts.

Zadeh, L., 1986. Test-score semantics as a basis for a computational approach to the representation of meaning. Literary and Linguistic Computing 1, 24–35.