A Unified Treatment of Uncertainties

Pei Wang
Center for Research on Concepts and Cognition
Indiana University
pwang@cogsci.indiana.edu

September 29, 1993

Abstract

"Uncertainty in artificial intelligence" is an active research field, where several approaches have been suggested and studied for dealing with various types of uncertainty. However, it's hard to rank the approaches in general, because each of them is usually aimed at a special application environment.

This paper begins by defining such an environment, then show why some existing approaches cannot be used in such a situation. Then a new approach, Non-Axiomatic Reasoning System, is introduced to work in the environment. The system is designed under the assumption that the system's knowledge and resources are usually insufficient to handle the tasks imposed by its environment. The system can consistently represent several types of uncertainty, and can carry out multiple operations on these uncertainties. Finally, the new approach is compared with the previous approaches in terms of uncertainty representation and interpretation.

1 The Problem

The central issue of this paper is *uncertainty in intelligent reasoning system.* Though it is too early to establish a universally accepted definition for intelligent reasoning system, I want to give a definition to make it clear that what type of system I'm talking about. Why such a definition is chosen is explained in [32].

By *reasoning system*, I mean an information processing system that has the following components:

a representation language which is defined by a formal grammar, and used for the internal representation of the system's knowledge;

a communication language which is also defined by a formal grammar (may be identical to the previous one, but not necessarily), and used for the communication between the system and the environment;

an interface which provides a mapping of the two formal languages, so that the system can accept knowledge from the environment, and answer questions according to its knowledge;

an inference engine which use some rules to match questions with knowledge, to generate conclusions from promises, and to derive subquestions from questions;

a memory which store the tasks to be processed, and the knowledge according to which the tasks are processed;

a control mechanism which is responsible for the choosing of premise(s) and inference rule(s) in each step of inference, and the maintaining of the memory;

an interpretation which (maybe loosely) relates the two formal languages to natural language, so to make the system's behavior understandable to human beings.

The *environment* of a reasoning system can either be human users or other computer systems. In the simplest case, the system accepts two types of *tasks* from its environment: *new knowledge* and *questions*. The system should provided *answers* to its environment for the questions, according to available knowledge. For new knowledge, no reply is required. Therefore, the history of communications between such a system and its environment can be exactly recorded by the system's *experience* and *responses*, where the former is a sequence of *input knowledge* and *input questions*, the latter is a sequence of *answers*, and each piece of them is a sentence of the communication language.

Intelligence is understood here as *the ability of working and adapting to the environment with insufficient knowledge and resources.* More concretely, to be an information processing system that works under the *Assumption of Insufficient Knowledge and Resources* means the system must be, at the same time,

a finite system — the system's computing power, as well as its working and storage space, is limited;

a real-time system — the tasks that the system has to process, that is, new knowledge and questions, can emerge at any time, and all questions have deadlines attached with them;

an ampliative system — the system not only can retrieve available knowledge and derive sound conclusions from it, but also can make defeasible hypotheses and guesses based on it when no certain conclusion can be drawn; and

an open system — no restriction is imposed on the relationship between old knowledge and new knowledge, as long as they are representable in the system's communication language.

Furthermore, to be an *adaptive system* (or *learning system*) means the system must also be

a self-organized system — the system can accommodate itself to new knowledge, and adjust its memory structure and mechanism to improve its time and space efficiency, under the assumption that future situations will be similar to past situations.

Therefore, the system need rules for (at least) three types of operation:

Comparison: To choose an answer among several competing ones;

Revision: To modify the truth value of a piece of knowledge, in the light of new evidence;

Inference: To derive a conclusion from a set of given premises.

In the following discussions, we will see that for an intelligent reasoning system defined as above (henceforth IRS), there are all kinds of uncertainties in the system, such as randomness, fuzziness, ignorance, imprecision, incompleteness, inconsistence, and so on ([23]). Therefore, uncertainty should be taken into consideration when the above operations are carried out.

In this paper, I'll concentrate on the *representation* and *interpretation* of uncertainty. By *representation*, I mean the "label" attached to each sentence in the representation and communication language, which indicates the type and/or degree of uncertainty the sentence has. By *interpretation*, I mean the "meaning" of the labels, or how they are related to the methods that human beings use in everyday life to represent uncertainty.

2 The existing approaches

If there is an existing approach can do the job, then we can simply implement it into IRS. Unfortunately, there is no such an approach. I'll demonstrate why it is the case in the following.

2.1 The non-numerical approaches

The existing approaches for uncertainty management can be divided into two classes: numerical and non-numerical. The basic difference between them is: the former attach one or several *numbers* to each piece of knowledge to represent the degree of uncertainty, while the latter doesn't ([1, 3]).

Though using numbers to express "degree of uncertainty" seems to be a natural idea, there are still sound objections against it: Human knowledge is usually represented in natural language, where uncertainty is usually expressed verbally. If we can directly represent and process uncertainty in this form, the processes and results may be easier to be understood by human beings. When uncertainty is represented numerically, some information about the type and origin of the uncertainty will be lost, and it may lead to a false sense of accuracy. Moreover, it is hard to find a natural mapping between the numbers and the verbal expressions in natural language ([14, 24, 27]).

I'll discuss two typical non-numerical approaches: Non-Monotonic Logics and Endorsement Theory.

There are several formal systems within the category of Non-Monotonic Logics. Though built differently, they share some opinions about human common-sense reasoning: with incomplete knowledge, some convention or *default rules* can (and should be) used to get tentative conclusions, which can be rejected by later acquired facts.

If Non-Monotonic Logics is implemented as an IRS, the following problems will appear:

1. The system can accept default rules from the environment, but cannot generalize them from available evidence.

2. The domain knowledge of the system can be divided into three types: (1) default rules, (2) facts, and (3) guesses, where only the last one is defeasible. In this way, the system is not completely *open*: evidence that conflict with current default rules and facts cannot be accepted.

3. When there are competing guesses generated by different default rules, there is no universal way to make the choice. This is the so-called 'Multiple extensions problem".

Reiter wrote in [18]: "Nonmonotonic reasoning is necessary precisely because the information associated with such settings requires that certain conventions be respected." Since IRS is defined as a system that open to all representable new knowledge, no such convention can be assumed here.

Reiter wrote in the same paper that "Nonmonotonic reasoning is intimately connected to the notion of prototypes in psychology and natural kinds in philosophy." However, for an (completely) open system, the notion of prototypes and natural kinds have to be defined and maintained according to quantitative information. Concretely, prototype or default are based on the "central tendency" ([19]), but with constantly coming evidence, whether a tendency is "central" is usually a matter of degree. On the other hand, all knowledge is revisible, but with different sensitivity or stability, which is also a matter of degree. It is possible to indicate these degrees by verbal labels, but such a system will be less general and less efficient than a numerical system.

Endosement Theory uses verbal labels to represent "reasons to believe or disbelieve uncertain propositions" ([24]). Its advantage is the ability to indicate *why* to believe, as well as *how much* to believe, therefore more information is preserved, compared with numerical approaches.

We meet a dilemma here. If the endorsements are interpreted as different degrees along the same semantic dimension, what we get is a "coarse numerical scale", which has finite different values to take. Beside its naturalness (since verbal labels are used), such an approach has few advantage over numerical approaches ([27]). On the other hand, if the endorsements are interpreted as along different semantic dimensions (as the case in Cohen's papers), there must be labels that cannot be compared or used together as premises to derive conclusions, since they don't have a common measurement. In the situations where our main purpose is to *record* uncertainty, such an approach may be appropriate, but it is inapplicable in IRS, where the system *have to* set up a common representation for uncertainty from different sources, so that to carry out the operations (comparison, revision, and inference) on them.

Therefore, a numerical measurement of uncertainty is *necessary* for the *representation* language of IRS, not because it is accurate, but it is uniform and simple, especially because IRS need to do *induction* and *abduction*. In these inferences, uncertainty emerge even if all the premises are certain, and the amount of evidence is a dominant factor for the uncertainty (see [32] for an example). For this reason, even if verbal labels of uncertainty is used in the communication language for the sake of naturalness, it is still desired to represent uncertainty numerically in the internal representation.

A numerical measure is not sufficient for uncertainty management in IRS since some operations are sensitive to the source of uncertainty. To solve such problems, some other method can be used as a *supplement*, rather than a *replacement*, of the numerical approach. For example, such a mechanism is described in [30] for detecting correlated evidence in revision.

Though non-numerical approach is unsuitable in IRS, their motivations need to be respected, that is, the numerical measurement should have a natural interpretation, so that the numbers can make sense to human beings. On the other hand, verbal expressions have advantages when used in communication. We'll see in the next sections that it is possible to absorb these ideas into a numerical representation of uncertainty.

2.2 The fuzzy approaches

There are two types of approaches that can be called "fuzzy": one is using *linguistic variables* to represent uncertainty, and the other is using *grade of membership* (or its variations, such as *possibility*, as in [37]) to do it. The former falls into the category of non-numerical approaches, so the previous discussion applies, that is, though such an approach may work well for the communication language, it is not good enough as an internal representation. Therefore, here we'll focus our attention on the latter approach: representing the uncertainty of a proposition by a real number in [0, 1]. To simplify our discussion, let's assume the proposition have the form "b is A", so the number is the grade of membership of b in A ([35]).

The problem of using such an approach in IRS is: fuzziness is not properly analyzed and interpreted in fuzzy logic (see [31] for a detailed discussion). According to Zadeh, membership functions are subjective and context-dependent, therefore, there is no general method to determine them by experiment or analysis ([36]).

One problem caused by this attitude to fuzziness is: the operations, typically the *max* and *min* functions for union and interjection of fuzzy sets, lack a cognitive justification.

Such an approach may works well in some circumstances, for example in control systems, where the context is relatively stable, so the designer can adjust the system by trying different membership functions and operators. However, this methodology cannot be applied in IRS, which is *general purpose*, and its context is dynamic changed, not completely predictable by the designer.

In IRS, all uncertainties, including fuzziness, come from the insufficiency of the system's knowledge and resources. Therefore, we need a interpretation that relate the grade of membership to the amount (or weight) of available evidence, and the interpretation should be domain independent and consistent with the interpretation of other types of uncertainty, such as randomness. Otherwise, the system cannot carry out its operations in situations where different types of uncertainty are involved ([31]).

2.3 The Bayesian approach

The Bayesian approach for uncertain reasoning is characterized by the following features ([16]):

1. probability is interpreted as degree of belief, based on available evidence;
2. current knowledge is represented by a (first-order, real-valued) probability distribution on a proposition space; and
3. new knowledge is learned by conditionalization.

This approach has a sound theoretical foundation and a wide application domain. However, it cannot be used by IRS, for the following reasons:

- As in the case of fuzziness, how the degree of belief is related to weight of evidence is unclear ([21]).
- Due to insufficient knowledge and resources, it is usually impossible for IRS to maintain a *consistent probability assignment* on its knowledge base ([9]).
- As discussed in [28], conditionalization cannot be referred as a general way to symmetrically combine evidence from different sources.
- With all the efforts to improve its efficiency, the resources expense of Bayesian approach is still pretty high for large knowledge bases.

The origin of the problems is: all the probability assignments in Bayesian approach are based on a common chunk of prior knowledge, which is implicitly indicated, therefore cannot be weighed against new evidence ([28]).

According to the requirement of IRS, what is lacking in Bayesian approach is:

1. the ability to represent *ignorance* (or its opposite: *confidence*), which indicate the sensitivity of the current belief to future evidence; [1]

2. the ability to base each piece of knowledge on *its own* evidence, rather than let all of them share a common foundation (only in this way, the system's belief can be revised *locally* and *incrementally*, when the system cannot afford the resource for global revision); and

3. the ability to get conclusion when the premises are evaluated by different probability distribution functions.

What follows from these is: at least two numbers should be attached to each proposition to represent its uncertainty ([28]).

2.4 The higher-order approaches

"Ignorance cannot be represented in a probability distribution" is an old argument against Bayesian approach, and a major motivation for alternative approaches.

Since by ignorance we mean that the assignment of a probability to a proposition is uncertain itself, it's natural to try probability theory once again by introducing *second-order probability* or *probability of probability*: if the proposition "Proposition A's probability is p" is uncertain, we can measure the uncertainty by assigning a probability q to it. To the original proposition A, q is its second-order probability.

Mathematically, this can been done, and have been done ([5, 6, 15]), but its interpretation and practical benefit is doubtful ([12, 16]). From the point of view of IRS, we can (at least) find the following two reasons that are against this approach:

- Though second order probability q can be used to represent the uncertainty in a probability assignment $Pr(A) = p$, it does not represent *ignorance*, since $q = 0$ means $Pr(A) \neq p$, rather than "$Pr(A)$ is completely unknown".

- Given insufficient knowledge, we cannot get for sure not only the first order probability, but also the second order probability. Therefore in theory there is an infinite regression.

There are other attempts to measure ignorance by introduce a number which is at a "higher level" in certain sense, but not a "probability of probability" as defined above. For examples, Yager's *credibility* ([34]) and Shafer's *reliability* ([22]) both evaluate the uncertainty of a probability assignment, where 0 is interpreted as "unknown", rather than "impossible". However, these approaches still lack clear interpretations and well-defined operations.

[1]The claim that ignorance can be derived from a probability distribution ([16]) is incorrect due to the confusion of the "explicit condition" and the "implicit condition" of a probability assignment, as discussed in [28].

2.5 The interval approach

Another intuitively appealing approach to measure ignorance is to use an *interval*, rather than a *point*, to represent the probability of a proposition, and interpret the interval as the lower bound and upper bound of the "objective probability" ([2, 8, 11, 13, 33]).

In this way, ignorance can be represented by the width of the interval. When the system know nothing about a proposition, the interval is [0, 1], that is, the objective probability can be anywhere; when the objective probability is known, the interval degenerates into a point. The operations on the interval can be got directly from probability theory, so the approach has a solid foundation.

On the other hand, this approach can be interpreted as a partially defined probability distribution, where the undefined part is restrainted by the defined part.

If we try to use this approach in IRS, we'll meet a problem similar to the problem with second-order probability: with insufficient knowledge, we even cannot determined the lower bound or upper bound of a unknown probability. For example, if we don't know whether a coin is biased, and all our knowledge about it is that at n tossings, m of them get heads ($m \leq n$). Of course we don't know the probability of head (defined as the limit of the frequency), but do we know the interval within which the probability stays? Since future evidence is infinite compared with known evidence, no such interval can be guaranteed, except the trivial [0, 1].

Of course we can *guess* such an interval, but then the reliability of the guess need to be indicated somehow — once again, we are facing an infinite regression.

Moreover, how can we know there is a "objective probability", or the frequency of head has a limit? In the environment of IRS, we don't know it. As a result, the probability-interval approach has difficulty to find a suitable interpretation, though it can have a good formal description.

2.6 The Dempster-Shafer approach

In Dempster-Shafer theory ([21]), the belief about a proposition is also represented as an interval $[Bel, Pl]$, but it isn't interpreted as the lower and upper bounds of an objective probability, as least according to Shafer ([22]). As "a mathematical theory of evidence", Dempster-Shafer theory is characterized by the using of Dempster's rule to combine degrees of belief when they are based on independent items of evidence.

Shafer claimed that probability is a special case of degree of belief, when the $[Bel, Pl]$ interval degenerated into a point ([21, 22]). In [29], I argued that such a claim is misleading. I'll briefly summarize my argument here:

Consider a simple *frame of discernment* $\Theta = \{H, \bar{H}\}$. If w^+ is the *weight of evidence* that support H, and w^- is the *weight of evidence* that support $\bar{H}$, then it follows from Dempster's rule that

$$Bel(\{H\}) = m(\{H\}) = \frac{e^{w^+} - 1}{e^{w^+} + e^{w^-} - 1}$$

$$Pl(\{H\}) = 1 - m(\{\bar{H}\}) = \frac{e^{w^+}}{e^{w^+} + e^{w^-} - 1}$$

which are derived in [21].

The above equations lead to two results when the $[Bel, Pl]$ interval degenerated into a point, that is, when the weight of total evidence $w^+ + w^-$ go to infinite:

1. If w^+ and w^- keep a certain *proportion*, say $p : (1-p)$, when they go to infinity, then we have

$$Bel(\{H\}) = Pl(\{H\}) = \begin{cases} 0 & \text{if } p < 1/2 \\ 1/2 & \text{if } p = 1/2 \\ 1 & \text{if } p > 1/2 \end{cases}$$

2. If w^+ and w^- keep a certain *difference*, say $w^- - w^+ = \Delta$, when they go to infinity, then we have

$$Bel(\{H\}) = Pl(\{H\}) = \frac{1}{1+e^{\Delta}}$$

These results indicate that when a $[Bel, Pl]$ interval degenerated into a point, we get a "probability" function only in the sense that $Bel(\{H\}) + Bel(\{\bar{H}\}) = 1$, but usually (unless the point happens to be 0, 1/2, or 1) $Bel(\{H\})$ doesn't equal to p, which is commonly accepted as H's "probability". In most cases, such a *Bayesian belief function* corresponds to a situation where the positive and negative evidence have a constant *difference*, which has no obvious practical usage ([21]).

Therefore, if we still relate the probability of a proposition to the frequency of its positive evidence or the limit of the frequency, then it is neither necessarily in $[Bel, Pl]$, nor does it go to the same limit with the interval.

Since the frequency (or proportion) of positive evidence is a very natural and simple way to indicate the uncertainty of a proposition, the above result shows a defect of Dempster-Shafer theory, because this information is hard to access, and even lost when the interval degenerated into a point. For a detailed discussion of the problem, see [29].

Beside the above problem, there are some other factors that make Dempster-Shafer theory unsuitable for IRS, such as

- Dempster-Shafer theory only works on *exclusive* and *exhaustive* sets of hypotheses. However, not all sets of competing hypotheses have the two properties, due to insufficient knowledge and resources.
- Dempster's rule need unrealistic assumptions about the independence of the evidence to be combined, at least under the message-coding interpretation ([26]).
- Dempster-Shafer theory lack well-developed inference rules ([1]). This is partly caused by the above problem, because Dempster-Shafer theory cannot be treated as a generalization of probability theory in all senses.
- The resource expense of Dempster-Shafer theory's implementation is very high.

2.7 Summary

After above discussions we can say that though the existing approaches have their advantage in their applicable domains, none of them can be used in IRS. There are three basic reasons:

1. IRS is *completely* based on the assumption of *insufficient knowledge and resources*, while all the previous discussed approaches are *partially* based on the assumption. They admit that some knowledge is unavailable, but insist the existence of some other knowledge, and none of them consider how to work under a variable time pressure.

2. The interpretation of the numbers (for numerical approaches), especially how they are related to weight of evidence, is unclear (such as in fuzzy approach), unrealistic (such as in probability-of-probability and probability-interval approaches), or unnatural (such as in Dempster-Shafer theory). As a result, it is hard for human beings to communicate with the system, or to justify the operations of the system.

3. None of the existing approaches can naturally handle randomness, fuzziness and ignorance in a consistent way, while this property is necessary for IRS.

Therefore, a new approach is required by IRS.

3 The NARS Approach

Non-Axiomatic Reasoning System, or NARS, is proposed as a way to build an intelligent reasoning system (IRS). Here, I'll concentrate on its representation and interpretation of uncertainty, and leave the other components (such as inference rules, memory structure, and control mechanism) to other papers. For more detailed and complete descriptions for NARS, see [32] and [30].

3.1 The cardinal form

As mentioned previously, NARS need to measure the *weights of (positive and negative) evidence* of a proposition. Such a task can be divided into three subtasks: to represent a proposition, to distinguish positive and negative evidence of a proposition, and to determine the unit of weight.

In NARS, each proposition has the form "$S \subset P$", where S is the *subject* of the proposition, and P is the *predicate*. Both of them are *terms*. "$\subset$" is an *inheritance relation* between two terms, and is reflexive and transitive. Due to insufficient knowledge and resources, such a relation is usually uncertain, in the senses that there may be counter-examples, and the current evaluation may be revised by future evidence.

As an idealization, a binary reflexive and transitive relation "$\sqsubset$" is defined as the limit of the "$\subset$" relation, when the relation has no and will not have any counter-example. The *extension* of a term X is defined as the set of terms $\{Y|Y \sqsubset X\}$, and The *intension* of a term X is defined as the set $\{Y|X \sqsubset Y\}$. Intuitively speaking, the extension of a term is its *specializations*, or subsets; the intension of a term is its *generalizations*, or supersets. (See [30] for a detailed description of the representation language.)

"$S \subset P$" is called a "inheritance relation" from S to P, since in its idealized case (represented by "$S \sqsubset P$") S inherits P's intension, and P inherits S's extension. As said previously, such a inheritance relation is usually uncertain, so it is necessary to measure the weight of evidence. According to the above definitions, if it is known that "$M \sqsubset S$" and "$M \sqsubset P$", or "$S \sqsubset M$" and "$P \sqsubset M$", then M can be naturally counted as a piece of *positive* evidence for "$S \subset P$", with a *unit weight*, since such a M is a common element in S's and P's extensions or intensions, therefore supports the judgment that "S inherits P's intension, and P inherits S's extension" (which is what formalized by "$S \subset P$") to a certain extent. On the contrary, if it is known that "$M \sqsubset S$" but "$M \sqsubset P$ is false", or "$P \sqsubset M$" but "$S \sqsubset M$ is false", then M can be naturally counted as a piece of *negative* evidence for "$S \subset P$", with a *unit weight*, since such a M is an element in S's extension but not in P's, or an element in P's intension but not in S's, therefore rejects the judgment that "S inherits P's intension, and P inherits S's extension" to a certain extent.

Therefore, if we write the extension of a term X as E_X, and the intension of X as I_X, then for a proposition "$S \subset P$", the weights of its positive and negative evidence can be defined in terms of the *cardinals* of the related sets:

$$w^+ = |E_S \wedge E_P| + |I_P \wedge I_S|$$

$$w^- = |E_S - E_P| + |I_P - I_S|$$

respectively, and the weight of *all relevant evidence* is defined as

$$w = w^+ + w^- = |E_S| + |I_P|$$

There are several points to be noticed in the definition:

1. Weight of evidence, defined in this way, is not used to be actually, directly measured for each proposition, but used to interpret the uncertainty and to provide a foundation for the operations on uncertainty.

2. All terms, as pieces of positive and negative evidences for an inheritance relation, are equally weighted.

3. What is counted as the cardinal of a set is not "how many terms are there" which has a objective sense, but "how many times a term has been known (by the system) to be there" which is system dependent.

4. Practically, a term can hardly be counted as a "ideal element" of an extension or intension, so its weight is usually a decimal, less than 1. Therefore, the cardinal of such a "fuzzy set" is not necessarily to be an integer.

3.2 The ratio form

Though the cardinal form of uncertainty is logically more basic, it is unnatural and inconvenient in many situations. We often prefer a "relative measurements", such as a real number in $[0, 1]$. It's easy to see that

$$f = \frac{w^+}{w}$$

will give us the "success frequency" of the inheritance relation between the two terms, according to the system's experience.

This measurement is natural and useful, but not enough: we still need the information about the absolute value of w^+ or w, so to manage the future revision of the frequency ([28]).

Can we find a natural way to represent this information in the form of a "relative measurements", or more specificly, as a *ratio*? An attractive idea is to define "the second-order probability". However, as discussed previously, it doesn't make sense to compare the amount of relevant past experience, represented by w, to the future experience, which is (potentially) infinite.

However, it makes perfect sense to talk about the "near future". What the system need to know, by keeping the information about w, is how sensitive a frequency will be to new evidence, then use the information to make choice among competing judgments. If we limit our attention to the same "constant future", we can keep such information in a *ratio* form.

Let's introduce a positive constant k, and say it indicates that by "near future", we mean "to test the inheritance relation for k more times", then there is a natural way to represent the system's *confidence* (indicated by c) about the frequency:

$$c = \frac{w}{w + k}$$

that is, as the ratio that the amount of "current relevant evidence" to the amount of "relevant evidence in the near future". Intuitively, 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 (represented by a bigger w), the more confident the system is about the frequency, since the effect of evidence that comes in the near future will be relatively smaller (we'll see how c actually works in the revision operation in [30]).

We can also naturally define *ignorance* as the complement of confidence by

$$i = 1 - c.$$

Confidence and ignorance, when defined like these, are measurement about the stability or sensitivity of the frequence, by consider its variability in the near future. By indicating the near future by a constant k, what I mean is not that NARS always requires the same level of stability for different propositions, but that it is a way to define a *unit* for the measurement of stability.

3.3 The interval form

Amazingly, there is a third way to represent and interpret a truth value in NARS: to represent it as an interval. Obviously, no matter what will happen in the *near future*, the "success frequency" will be in the interval

$$[\frac{w^+}{w + k}, \frac{w^+ + k}{w + k}]$$

after the constant period indicated by k. This is because the current frequency is $\frac{w^+}{w}$, so in the "best" case, when all evidence in the near future is positive, the new frequency will be $\frac{w^+ + k}{w + k}$; in the "worst" case, when all evidence in the near future is negative, the new frequency will be $\frac{w^+}{w+k}$.

Let's write the interval as $[a, z]$, and simply call the two values as the *lower bound* and *upper bound* of the frequency, but keep in mind that they respect to a constant *near future*, rather then an *infinite future*. Such an interval have some interesting properties:

1. The width of the interval is exactly the ignorance as defined above, that is, $z - a = i = 1 - c$.

2. The frequency f divide the $[a, z]$ interval into the same proportion as it divide the $[0, 1]$ interval, which is the proportion between the weights of the positive and the negative evidence, that is, $f - a : z - f = f : 1 - f = w^+ : w^-$.

3.4 Summary

Now we have three functionally identical ways to represent the uncertainty in a proposition in NARS ([30]):

1. as a pair of *cardinals* $\{w^+, w\}$, where w^+ is a non-negative real number, w is a positive real number, and $w \geq w^+$;

2. as a pair of *ratios* $<f,\, c>$, where $f \in [0,1]$, and $c \in (0,1)$; or

3. as an interval $[a,\, z]$, where $0 \leq a < z \leq 1$, and $1 > z - a$.

Beyond these valid truth values, there are two limitation points useful for the defining of the inference rules:

Null evidence: This is represented by $w = 0$, or $c = 0$, or $z - a = 1$. It means that the system actually know nothing at all about the inheritance relation;

Total evidence: This is represented by $w \to \infty$, or $c = 1$, or $z = a$. It means that the system already know everything about the statement — no future modification of the truth value is possible.

These are the one-to-one mappings among the three forms:

	$\{w^+, w\}$	$<f,\, c>$	$[a,\, z]$
$\{w^+, w\}$		$w^+ = k\frac{fc}{1-c}$ $w = k\frac{c}{1-c}$	$w^+ = k\frac{a}{z-a}$ $w = k\frac{1-(z-a)}{z-a}$
$<f,\, c>$	$f = \frac{w^+}{w}$ $c = \frac{w}{w+k}$		$f = \frac{a}{1-(z-a)}$ $c = 1 - (z - a)$
$[a, z]$	$a = \frac{w^+}{w+k}$ $z = \frac{w^+ + k}{w+k}$	$a = fc$ $z = 1 - c(1 - f)$	

Since the three form is functionally identical, we can use any of them in the internal representation language. However, this doesn't mean that the other two forms are redundant and useless. To have different, but closely related forms and interpretations for truth value has many advantages:

- It gives us a better understanding about what the truth value really means in NARS, since we can explain it in different ways. The mappings also tell us the interesting relations among the various ways of uncertainty measurements.

- As we'll see in the next section, it makes us more convenient to compare this approach with other approaches.

- It makes the designing of inference rules easier. For each rule, there are functions calculating the truth value of the conclusion(s) from the truth values of the premises, and different functions correspond to different rules. As shown in [30], for some rules it is easier to choose a function if we treat the truth values as cardinals, while for other rules, we may prefer to treat truth values as ratios or intervals.

The above results also provides a user-friendly interface: if the environment of the system is human users, the uncertainty of a statement can be represented in different forms, such as "I've tested it w times, and in w^+ of them it was true", "Its past success frequency is f, and its confidence is c", or "I'm sure that its success frequency with remain in the interval $[a,\, z]$ in the near future". Using the mappings in the above table, we can maintain an unique truth value form as internal representation, and translate the other two into it in the interface.

More than that, it is possible for NARS to support other ways to represent uncertainty in the communication language:

Single number: We can use a single decimal to carry the information about frequency and ignorance. For example, 0.9 can be translated into the interval $[0.85, 0.95]$, and 0.90 can be translated into the interval $[0.895, 0.905]$.

Linguistic variables: We can use a set of verbal labels to approximately represent frequence and ignorance. In the simplest cases, if there are n labels in the set, the [0,1] interval is evenly divided into n sub-intervals, each for one of the labels. For example, the label set {false, more or less false, borderline, more or less true, true} can be translated into [0, 0.2], [0.2, 0.4], [0.4, 0.6], [0.6, 0.8], and [0.8, 1], respectively (similar methods are discussed in [27]). [2] It is also possible to determine the meanings of everyday verbal uncertainty expressions by psychological experiments ([17, 27]).

These two forms don't have one-to-one mappings with the three forms defined above, but they are good for communication in the situations where naturalness and simplicity are weighed more than accuracy, or to "convey the vagueness, or softness, of one's opinions" ([27]). As a result, though within the system the uncertainty of propositions is consistently represented by one form, it can be represented by (at least) five different (but related) forms in the communication language.

It is also possible to extend the approach of uncertainty representation and interpretation to other formal language, where the propositions are not in the form "$S \subset P$". What is necessary is that the positive and negative evidence of a proposition can be consistently distinguished and naturally measured.

4 Relationship to Other Approaches

Now let's briefly compare the representation and interpretation of uncertainty in NARS with the other approaches mentioned previously, to see their similarities and differences.

4.1 The non-numerical approaches

Though NARS uses a numerical representation for uncertainty in its (internal) representation language, it also allow certain verbal descriptions of uncertainty in its (external) communication language, as mentioned in the previous section.

NARS also have some functions similar to non-monotonic logics: it can make guesses when the knowledge is insufficient, and the guesses are based on the "typical" or "normal" situations, according to the system's knowledge. When such guesses conflict with new evidence, they can be modified, or even rejected.

As a numerical approach, NARS have more types of operations on uncertainty than non-monotonic logics, for example, NARS can generate hypotheses from evidences by induction and abduction, and all of its knowledge is revisible.

4.2 The fuzzy approaches

By the way the uncertainty of a proposition is represented and interpreted, NARS suggest a new interpretation of fuzziness (see [31] for a detailed discussion). Here I only discuss how two simple

[2] For a comparison, in fuzzy logic the set may be translated into {0, 0.25, 0.5, 0.75, 1} as in [4], or into five 4-tuples as in [3].

types of fuzziness are interpreted and represented in NARS.

The first type of fuzziness mainly happens with adjectives and adverbs, and appears in sentences with the pattern "A is a R C", such as "John is a young man" and "Tweety is a small bird", where C (man, bird) is a class of objects, A (John, Tweety) is an object in C, and R (young, small) is an adjective those comparative form "R-er than" (younger than, smaller than) is a binary relation on C, which is asymmetrical, transitive, and non-fuzzy.

In such a case, "R C" (young man, small bird) is a fuzzy concept, because the information is given by comparing an object to a reference class. Under such a situation, it is not surprise to see that membership is a matter of degree, since "R C" means "R-er than the other Cs", which usually has both positive evidence and negative evidence, depending on which other element in C is compared to.

In NARS, this "grade of membership" can be easily measured by the frequency of positive evidence, which provides A's relative ranking in C with respect to the relation "R-er than", or R_t:

$$\mu_{R\_C}(A) = \frac{|(\{A\} \times C) \cap R_t|}{|C - \{A\}|}$$

(it's not necessary to compare A to itself). Now $\mu_{R\_C}(A) = 1$ means that A is the "R-est C" (youngest man, smallest bird); $\mu_{R\_C}(A) = 0$ means that A is the "un-R-est C" (oldest man, biggest bird).

The second type fuzziness mainly happens with nouns and verbs, where the membership is a *similarity* between an instance to be judged and a prototype or a known instance ([20]), so membership measurement is reduced to similarity measurement.

In NARS, the inheritance relation is exactly what usually referred as *asymmetric similarity*, that is, "S inherit P's property". The grade of membership of an element of S in P is f, the frequency of positive evidence, as defined before. Now $f = 1$ means that each elements of S inherits all the properties of P, so they are typical P; $f = 0$ means that each elements of S inherits no properties of P, so they are by no means P.

We can define *symmetric similarity* as two inheritance relations, that is, to understand "S and P are similar" as "S inherit P's property, and P inherits S's properties". The *degree of similarity* also be naturally measured by the frequency of positive evidence

$$\frac{|I_S \cap I_P|}{|I_S \cup I_P|}$$

where I_S (I_P) is S's (P's) property set, or its *intension*, as previously defined.

Of course, in all these situations, the confidence of the property also need to be calculated, to make the frequencies revisible in light of future evidence.

Compared with the various fuzzy logics, the process of fuzziness in NARS has the following characteristics:

1. The membership function is related to the weight of evidence, and it is explicitly defined in each case that what is counted as positive or negative evidence.

2. Fuzziness and randomness are both represented by the frequency of positive evidence, and processed in a unified way. The difference is: the former is about the inheritance of intensions, while the latter is about the inheritance of extensions.

3. A confidence measurement is used to maintain the revision of grade of membership.

4.3 The Bayesian approach

Though the NARS approach of uncertainty management is closely related to probability theory, there are still several important differences.

The *frequency* measured in NARS is about "psychological events" (whether two terms share their instances/properties) of the system, so it is *empirical* in the sense that it is determined by the experienced frequency, it is *subjective* in the sense that it depends on the system's individualized environment and internal state, and it is *logical* in the sense that it is interpreted as a logical relation between a judgment and available evidence. Therefore, it is related to all the three major interpretations of probability ([7, 10]), but identical to none of them.

NARS processes both extensional factors and intensional factors by which a proposition become uncertain, while traditionally probability theory is used for extensions only, that is, the probability of $S \subset P$ is usually measured by $\frac{|S \cap P|}{S}$.

The *confidence* measurement is introduces to support the revision of frequency ([28]).

In Bayesian approach, the probability assignments on a proposition space are based on a common prior knowledge base, so all beliefs must be *consistent*, and all operations must be take the entire content of the knowledge base into consideration. Pearl calls such approaches "intensional", and argued that they are more appropriate than the "extensional" approaches (the words "extension" and "intension" are used differently in Pearl's book and in this paper), where operations are *local* ([16]), that is, what conclusion can be drawn from a set of premises is independent to what other beliefs are hold by the system.

NARS is "extensional" in Pearl's sense, because in it each proposition's uncertainty is measured on its own, and all operations only involve a small part of the knowledge base. As a result, there are usually conflicting beliefs in the system's knowledge base, and the system may make mistakes due to the incomplete record of each belief's sources of evidence (see [30] for a detailed discussion). However, these problems is caused by the insufficiency of knowledge and resources, so are inevitable for IRS (as well as for human beings). In contrary, the "intensional" approach, such as Bayesian approaches, are applicable only in situations where the system's knowledge is sufficient on the relations between evidence from different sources, and the system's resources are sufficient for globe updating when new evidence comes ([28]).

4.4 The higher-order approaches

As discussed previously, though the confidence c of a proposition is in $[0, 1]$, can be measured as a ratio, and *is* at a higher level than the frequency of positive evidence f, in the sense that c indicates f's stability, c cannot be interpreted as a second-order probability h in the sense that it is the probability of the judgment "the inheritance relation's (real, or objective) probability is f", and cannot be processed in that way according to probability theory.

When $c = 1$, it means the same with $h = 1$, that is, f cannot be changed by future evidence. But $c = 0$ means f is undefined, while $h = 0$ means f is not the "true" probability.

The more fundamental difference is: by defining a second order probability, it is assumed that there is a "true" or "objective" probability, and its relation with the current "first order probability" is partially known. Such an assumption, though valid at some situations, is conflict with the assumption of insufficient knowledge, so cannot be accepted by IRS. On the contrary, *confidence* is totally defined on the system's experience, without any assumption about "the state of affairs". A proposition with a high confidence is more stable, or harder to be influenced by new evidence, but it doesn't mean that it is "truer" in an objective sense.

In this way, NARS also avoids the infinite regression mentioned previously: if the second order probability is uncertain, a third order estimation is necessary, at least in theory. In NARS, both f and c are *certain* in the sense that they are measurements on the system's experience, rather than estimations of unknown quantities existing outside the system. Therefore, no further quantities is necessary to represent the uncertainty of f and c.

As mentioned previously, the intuition behind *confidence* is similar to whose of Yager's *credibility* ([34]) and Shafer's *reliability* ([22]). What differs *confidence* from them is its explicitly defined relations with the other uncertainty measurement, such as weight of evidence, ignorance, and the range the frequency will be in the near future. With these relations, all these measurements form a consistent foundation for the NARS project.

4.5 The interval approach

The interval form of truth value in NARS shares similar intuitions with the "probability interval" approaches ([2, 11, 33]). For example, *ignorance* can be represented by the *width* of the interval. However, in NARS the interval is defined as the range the frequency will be in the *near* future, rather than in the *infinite* future. In this way, some theoretical problem can be solved.

As in the case of the second order probability, under the assumption of insufficient knowledge, it is invalid to assume there is an objective probability of a proposition, or the frequency has a limit. Therefore, it is nonsense to talk about an interval where the probability will be. However, by only talking about a constant future, such an interval can make perfect sense.

Interpreted in this way, the $[a, z]$ interval is processed differently from those interpreted as "lower/upper bound of (objective) probability". For example, during revision, two intervals that have no common sub-interval can still be combined, rather then treated as an inconsistency ([8]).

4.6 The Dempster-Shafer approach

The $[a, z]$ interval is similar to the $[Bel, Pl]$ interval of Dempster-Shafer theory in the sense that both can represent the *support* a proposition get from available evidence, as well as the system's ignorance about the proposition. On the other hand, both approaches try to make the point-like probability assignment a special case of the interval.

However, as demonstrated previously, in Dempster-Shafer theory the weight of evidence and the $[Bel, Pl]$ interval have such a relation that the information about the frequency of positive evidence is hard to access, and even be lost when the interval degenerated into a point. On the contrary, this information is explicitly reserved in the $[a, z]$ interval, which is defined as the range of the frequency in the near future, and will converge to the limit of the frequency when such a limit happen to exist.

5 Discussions

Compared with the other approaches, the representation and interpretation of uncertainty in NARS have the following characteristics:

1. It can satisfy the requirement of IRS, that is, the approach can be applied to a reasoning system where knowledge and resources are constantly insufficient to deal with the tasks provided by the environment.

2. It combines various measurements of different types of uncertainty into a unified framework, and provides them with natural and consistent interpretations.

3. It provides a consistent foundation for the uncertainty calculus which includes several kinds of operations, such as comparison, revision, and inference (deduction, induction, abduction, and so on).

NARS process multiple types of uncertainty in a unified way, not because they are not different (on the contrary, they *are* different, see [23]), but because the system *have to* find a common measurement at a more abstract level, otherwise the system would fail to satisfy the requirements of the environment. I hope this paper can show that such a common measurement is possible. In NARS, different types of uncertainty can still be distinguished in extreme cases, but it is not always possible or easy.

It is not claimed that this approach is always better than the others, but that it is better in the environment defined as IRS in the first section of the paper. For why such an environment is important and interesting from the point of view of artificial intelligence and cognitive science, see [32].

This approach have many interesting properties, for example, it satisfies most requirements in Bonissone's list of desiderata ([2]). On the other hand, the system may make mistakes, but these mistakes are caused by the insufficiency of knowledge and resources, so they are "reasonable errors" in the sense that they are similar to human mistakes under the same situation (for example, some phenomena happened in NARS are like what Tversky and Kahneman discussed in [25] as "heuristics and biases"). Therefore, NARS is not only proposed as an engineering model for solving certain practical problems under certain situations, but also as a cognitive model to explain intelligence.

Acknowledgement

This work is supported by a research assistantship from Center for Research on Concepts and Cognition, Indiana University.

References

[1] R. Bhatnagar and L. Kanal. Handling uncertain information. In L. Kanal and J. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 3–26. North-Holland, Amsterdam, 1986.

[2] P. Bonissone. Summarizing and propagating uncertain information with triangular norms. *International Journal of Approximate Reasoning*, 1:71–101, 1987.

[3] P. Bonissone and K. Decker. Selecting uncertain calculi and granularity. In L. Kanal and J. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 217–247. North-Holland, Amsterdam, 1986.

[4] D. Dubois and H. Prade. *Fuzzy Sets and Systems*. Academic Press, New York, 1980.

[5] R. Fung and C. Chong. Metaprobability and Dempster-Shafer in evidential reasoning. In L. Kanal and J. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 295–302. North-Holland, Amsterdam, 1986.

[6] H. Gaifman. A theory of higher order probabilities. In J. Halpern, editor, *Theoretical Aspects of Reasoning about Knowledge*, pages 275–292. Morgan Kaufmann, Los Altos, California, 1986.

[7] I. Good. *The Estimation of Probabilities.* The MIT Press, Cambridge, Massachusetts, 1965.

[8] B. Grosof. An inequality paradigm for probabilistic knowledge. In L. Kanal and J. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 259–275. North-Holland, Amsterdam, 1986.

[9] G. Harman. *Change in View.* The MIT Press, Cambridge, Massachusetts, 1986.

[10] H. Kyburg. *The Logical Foundations of Statistical Inference.* D. Reidel Publishing Company, Boston, 1974.

[11] H. Kyburg. Bayesian and non-Bayesian evidential updating. *Artificial Intelligence*, 31:271–293, 1987.

[12] H. Kyburg. Higher order probabilities. In L. Kanal, T. Levitt, and J. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*, pages 15–22. North-Holland, Amsterdam, 1989.

[13] R. Loui. Interval-based decisions for reasoning systems. In L. Kanal and J. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 459–472. North-Holland, Amsterdam, 1986.

[14] J. McCarthy and P. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer and D. Michie, editors, *Machine Intelligence 4*, pages 463–502. Edinburgh University Press, Edinburgh, 1969.

[15] G. Paaß. Second order probabilities for uncertain and conflicting evidence. In P. Bonissone, M. Henrion, L. Kanal, and J. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*, pages 447–456. North-Holland, Amsterdam, 1991.

[16] J. Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann Publishers, San Mateo, California, 1988.

[17] R. Reagan, F. Mosteller, and C. Youtz. Quantitative meanings of verbal probability expressions. *Journal of Applied Psychology*, 74:433–442, 1989.

[18] R. Reiter. Nonmonotonic reasoning. *Annual Review of Computer Science*, 2:147–186, 1987.

[19] E. Rosch. On the internal structure of perceptual and semantic categories. In T. Moore, editor, *Cognitive Development and the Acquisition of Language*, pages 111–144. Academic Press, New York, 1973.

[20] E. Rosch and C. Mervis. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7:573–605, 1975.

[21] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, New Jersey, 1976.

[22] G. Shafer and R. Srivastava. The Bayesian and belief-function formalisms: A general perspective for auditing. *Auditing: A Journal of practice and Theory*, 9, 1990.

[23] P. Smets. Varieties of ignorance and the need for well-founded theories. *Information Sciences*, 57-58:135–144, 1991.

[24] M. Sullivan and P. Cohen. An endorsement-based plan recognition program. In *Proceedings of the National Conference of Artificial Intelligence*, pages 475–479. 1985.

[25] A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185:1124–1131, 1974.

[26] F. Voorbraak. On the justification of Dempster's rule of combination. *Artificial Intelligence*, 48:171–197, 1991.

[27] T. Wallsten, D. Budescu, and R. Zwick. Comparing the calibration and coherence of numerical and verbal probability judgments. *Management Science*, 39:176–190, 1993.

[28] P. Wang. Belief revision in probability theory. In *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence*, pages 519–526, Washington, DC, 1993. Morgan Kaufmann Publishers, San Mateo, California.

[29] P. Wang. A defect in Dempster-Shafer theory. Technical Report 85, Center for Research on Concepts and Cognition, Indiana University, Bloomington, Indiana, 1993.

[30] P. Wang. From inheritance relation to non-axiomatic logic. Technical Report 84, Center for Research on Concepts and Cognition, Indiana University, Bloomington, Indiana, 1993.

[31] P. Wang. The interpretation of fuzziness. Technical Report 86, Center for Research on Concepts and Cognition, Indiana University, Bloomington, Indiana, 1993.

[32] P. Wang. Non-axiomatic reasoning system (version 2.2). Technical Report 75, Center for Research on Concepts and Cognition, Indiana University, Bloomington, Indiana, 1993.

[33] K. Weichselberger and S. Pöhlmann. *A Methodology for Uncertainty in Knowledge-Based Systems*. Springer-Verlag, Berlin, 1990.

[34] R. Yager. Credibility discounting in the theory of approximate reasoning. In P. Bonissone, M. Henrion, L. Kanal, and J. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*, pages 299–310. North-Holland, Amsterdam, 1991.

[35] L. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[36] L. Zadeh. A fussy-set-theoretic interpretation of linguistic hedges. *Journal of Cybernetics*, 2:4–34, 1972.

[37] L. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and System*, 1:3–28, 1978.