A New Approach for Induction: From a Non-Axiomatic Logical Point of View

Pei Wang
Center for Research on Concepts and Cognition
Indiana University
510 N. Fess, Bloomington, IN 47408
pwang@cogsci.indiana.edu

September 21, 1995

Abstract

Non-Axiomatic Reasoning System (NARS) is designed to be a general-purpose intelligent reasoning system, which is adaptive and works under insufficient knowledge and resources. This paper focuses on the components of NARS that contribute to the system's induction capacity, and shows how the traditional problems in induction are addressed by the system. The NARS approach of induction uses an term-oriented formal language with an experience-grounded semantics that consistently interprets various types of uncertainty. An induction rule generates conclusions from common instance of terms, and a revision rule combines evidence from different sources. In NARS, induction and other types of inference, such as deduction and abduction, are based on the same semantic foundation, and they cooperate in inference activities of the system. The system's control mechanism makes knowledge-driven, context-dependent inference possible.

1 Introduction

The term "induction" is usually used to denote the inference that derives *general* knowledge from *specific* knowledge. There are some people who call all non-deductive inferences "induction", but in this way the category includes too many heterogeneous instances to be studied fruitfully.

There are three major academic traditions in the study of induction. The philosophical/logical study concentrates on the formalization and justification of induction; the psychological study concentrates on the description and explanation of induction in the human mind; and the computational study concentrates on the implementation of induction in computer systems.

Though all these studies address induction in the above board sense, the precise ways in which "induction" is formulated are quite different. Consequently, the researchers actually work on different, though closely related, problems.

We can distinguish three kinds of formal model of induction, according to the unit of knowledge focused in the study: (1) concept, (2) declarative sentence, and (3) procedural sentence.

In the first kind of model, "induction" is defined as the procedure that takes descriptions of instances of a concept (as input), and generates a description of the concept (as output). Many machine-learning projects use the term "induction" in this sense (Michalski, 1983; Quinlan, 1986).

The second kind of model defines "induction" as descriptive generalization, whose results are sentences in a formal (declarative) language. Most philosophical/logical study and some AI study on induction belong to this category (Cohen, 1989; Kyburg, 1970; Michalski, 1993).

In the third kind of model, "induction" is defined as the process from specific instances to learn general cognitive skills. Such a treatment of induction can be found in (Holland et al., 1986).

In this paper, we will focus our attention to the second kind of model. We will first describe the related theories and introduce a new approach. Then, we will discuss how the new approach deals with the problems related to induction. Because our study is about *normative* models of induction, we will compare our approach with other related works in logic, philosophy and artificial intelligence, but ignore the psychological literatures on descriptive models of induction.

2 Background

Though Aristotle mentioned induction as the method by which general primary premises can be obtained, he did not develop a theory for this type of inference, as he did for deduction. It was Bacon who for the first time proposed a systematical inductive method, with the hope that it could provide a general methodology for empirical science (Cohen, 1989).

However, such an approach was seriously challenged by Hume, who argued that the inferences that extend past experience to future situations cannot have a logical justification (Hume, 1748). After Hume, most philosophical and logical work on induction are about the justification of the process. The mainstream approach is to use probability theory, with the hope that though inductive conclusions cannot be absolutely true, they can have certain probabilities (Carnap, 1950).

In recent years, the study of induction has been enriched by AI researchers. With computer systems as tools and platform, different formalizations and algorithms are proposed and tested.

As mentioned previously, in this paper "induction" is used for the inference in which general declarative knowledge is derived from, or confirmed by, specific knowledge. In terms of the formal language used, we can further divide the existing approaches in this domain into three "families".

The first family uses propositional logic and probability theory. Let us say that S is a proposition space and P is a probability distribution function on it. Induction is defined in this situation as the operation of determining $P(H|E)$, where H is a hypothesis and E is available evidence, and both belong to S. The inference — or more precisely, calculation — is carried out according to probability theory in general, and Bayes's theorem in particular. This family is the mainstream of the philosophical and logical tradition of induction study (Keynes, 1921; Carnap, 1950; Good, 1983), and it has been inherited by the Bayesian school in AI (Korb, 1995; Pearl, 1988).

The second family uses first-order predicate logic. Let us say that B is the background knowledge of the system, and E is available evidence (both B and E are sets of statements in first-order predicate logic). Induction is defined in this situation as the operation of finding a statement H that implies E and is also consistent with B. Because the inference from H and B to E is deduction, induction thus defined, as the inference from E and B to H, is often referred to as "reverse deduction". This family is very influential in machine learning (Michalski, 1993).

The third family uses term logic. This kind of logic, exemplified by Aristotle's system, is characterized by the use of subject–predicate sentence and syllogistic rules. Though Aristotle discussed induction briefly in his work (Aristotle, 1989), it was Peirce who first defined different types of inference in term logic, roughly in the following manner (Peirce, 1931):

deduction	**induction**	**abduction**
$M \subset P$	$M \subset P$	$P \subset M$
$S \subset M$	$M \subset S$	$S \subset M$
————	————	————
$S \subset P$	$S \subset P$	$S \subset P$

One interesting fact is that though Peirce's distinction of deduction, induction, and abduction is widely accepted, his formalization in term logic is seldom followed. Instead, the above definition is rephrased within the frame of first-order predicate logic (Michalski, 1993). We will see the subtle difference between these two formalizations later.

NARS, the new approach of induction that will be discussed in this paper, belongs to the term-logic family. NARS stands for *Non-Axiomatic Reasoning System*. It is a general-purpose reasoning system, which accepts knowledge provided by the user in a formal language, and answers questions according to available knowledge and various inference rules (Wang, 1994a; Wang, 1995b).

What distinguishes NARS from other reasoning systems is that it is designed to be *adaptive under insufficient knowledge and resources.*

Insufficient knowledge and resources means that the system works under the following restrictions:

Finite: The system has a constant information-processing capacity.

Real-time: The questions that the system need to answer have various time requirements attached.

Open: No constraints are put on the knowledge and questions that the system can accept, as long as they are expressible in the formal language.

To *adapt* means that the system learns from its experiences. It answers questions and adjusts its internal structure to improve its resource efficiency, under the assumption that future situations will be similar to past situations.

It follows from the above specifications that the ability of induction is necessary for NARS. On one hand, due to insufficient knowledge, the system needs to extend its previous knowledge to novel questions. On the other hand, due to insufficient resources, the system needs to compress its knowledge by generalization, so as to use its time–space resources more efficiently.

In the following, we only address aspects of NARS that are directly related to induction. For more comprehensive descriptions of the project, see (Wang, 1994a; Wang, 1994b; Wang, 1995b).

3 How to Represent Inductive Conclusions?

NARS uses a term logic, whose sentences (including all premises and conclusions in induction) are all in the "subject–predicate" format, as in the logics of Aristotle and Peirce. Formally, a statement in NARS has the following form:

$$S \subset P$$

where S is the subject term, and P is the predicate term. In the simplest situation, terms are just identifiers, or words, without any internal structure.

The relation "$\subset$" is an *inheritance relation*, which is defined in NARS by two properties: *reflexivity* and *transitivity*. Therefore, in ideal situations, we have

$$X \subset X$$

$$\{X \subset Y,\ Y \subset Z\} \vdash X \subset Z$$

where X, Y, and Z are arbitrary terms.

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 knows "$S \subset P$" for sure, then S can substitute P in sentences of the form "$P \subset X$", and P can substitute S in sentences of the form "$X \subset S$", where X is an arbitrary term. The other way around, if all X that satisfy "$X \subset S$" also satisfy "$X \subset P$", and all X that satisfy "$P \subset X$" also satisfy "$S \subset X$", then we have "$S \subset P$" (Wang, 1994a). In English, "$S \subset P$" roughly corresponds to "S is P", if we ignore the singular/plural distinction.

As mentioned previously, NARS is an adaptive system, and always open to new knowledge, meaning that all judgments the system makes are based on its experience. Consequently, whether "$S \subset P$" is true (i.e., whether there is an inheritance relation from S to P) is determined according to whether the system has experienced such a relation. In the simplest situation, the system's experience is the stream of input knowledge provided by the user, up to the current moment.

Such an experience-grounded semantics is fundamentally different from traditional model-theoretic semantics. In NARS, the truth value of a statement is not judged according to a constant set of axioms, and it may change as new evidence become available. What a truth value measures is the extent to which the statement is supported by available evidence, rather than the extent to which the statement is a matter of fact. Therefore, a truth value also indicates the system's degree of belief, or uncertainty, on the statement (Wang, 1994a; Wang, 1995a). We can see this more clearly later.

Truth value and degree of belief are usually treated as different properties of a statement — the former is objective and constant, while the latter is subjective and revisable. However, if we concern about what is true to a system which has insufficient knowledge and resources, we can see that it cannot judge truth without consulting its experience.

To decide truth according to available evidence and according to axioms are basically different. In the former situation, no decision is final in the sense that it cannot be revised by future evidence. Each piece of evidence, either affirmative or rejective, contributes to a certain extent to the evaluation of truth value. Therefore truth value is always a matter of degree in a system like NARS.

This opinion is against a well-known conclusion proposed by Popper. He argues that there is an asymmetry between verifiability and falsifiability — "a positive decision can only temporarily support the theory, for subsequent negative decisions may always overthrow it" (Popper, 1959).

The crucial point here is: what is the content of a general statement, or, in Popper's words, a theory?

According to our opinion, "Ravens are black" is a general statement, for which a black raven is a piece of positive (affirmative) evidence, and a non-black (e.g., white) raven is a piece of negative (rejective) evidence — the former verify an inheritance relation "$raven \subset$ *black-thing*" to a certain extent, while the latter falsify it, also to a certain extent. When we say that "All ravens are black", it means that according to our experience, the inheritance relation between the two terms only has positive evidence, but no negative evidence. In this case, the truth value of the statement is still a matter of degree, determined by the amount of available evidence.

What Popper refers to as theory are *universal statements*. Accordingly, when we say "All ravens are black", we means that all ravens in the whole universe, known or unknown, are black. Such a statement can only be true or false, and there is no middle ground (if we ignore the fuzziness of the terms). We know the statement is false as soon as we find a non-black raven, but we need to exhaust all ravens in the universe to know it is true.

Such a formalization of inductive conclusions is shared by the Baconian tradition of induction (Cohen, 1989). According to an approach proposed by Cohen, induction is a sequence of tests with increasing complexity, and the (Baconian) probability of a hypothesis indicates how many tests the hypothesis passed in the process.

If we accept the above definition of scientific theory, all conclusions of Popper and Cohen follow logically. However, why should we accept the definition? As a matter of fact, many empirical scientific theories have counterexamples, and we do not throw them away (Kuhn, 1970). It is even more obvious when we consider our common-sense knowledge. A general statement like "Ravens are black" works well as our guide of life, even when we know that it has counterexamples. Such a statement can be applied to predict new situations, though its truth value is determined by past experience. We do hope to establish theories that has no known counterexamples, but it does not mean that theories with known counterexamples cannot be used for various practical purposes. Only in mathematics, where truth values are determined according to fixed axioms, universal statements become available.

The above argument also serves as a criticism to the AI induction projects within the framework of binary logic (Korb, 1995). To define induction as "finding a pattern to fit *all* data" makes it a luxury that can only be enjoyed in a laboratory. Though such a paradigm can produce research results, these results are hardly extendable to practical situations. Also, this over-idealization makes the process fundamentally different from the generalizations happening in the human mind. It is even not appropriate to justify this approach as "a preliminary step toward more complex studies", because when giving up the idea that "an inductive conclusion can be falsified once for all", the situation will become so different that the previous results are hardly useful at all.

In summary, inductive conclusions, as other knowledge, are represented in NARS as inheritance relations between terms. Their truth values are not binary — either "true" or "false", but are indicated quantitatively, according to the experience of the system.

4 How to Define Truth Value?

Because in NARS truth values are determined by available evidence, we need to first precisely define what is counted as evidence and how evidence is quantitatively measured.

Though it is natural to say that a black raven is a piece of positive evidence for "Ravens are black", and a white raven is its negative evidence, Hempel points out that the concept of positive/negative evidence cannot be easily defined in the first-order (predicate-oriented) language in general (Hempel, 1943). Let us suppose that "Ravens are black" is formulated as $(\forall x)(Raven(x) \rightarrow Black\text{-}thing(x))$, and that a piece of positive evidence is a constant that when substituted into the variable x makes both the condition and the conclusion true. Consequently, a green shirt will also be counted as a piece of positive evidence for the sentence, because it confirms the "logically equivalent" sentence $(\forall x)(\neg Black\text{-}thing(x) \rightarrow \neg Raven(x))$. Such a result is highly counterintuitive, and may cause many problems (for example, a green shirt is also a piece of positive evidence for "Ravens are white", for exactly the same reason).

Here we will not discuss the various solutions proposed for this paradox. Almost all of these attempts are still within the framework of first-order predicate logic, whereas in the following we can see that the problem does not appear in term logics like NARS.

From the definition of inheritance relation introduced previously, we see that if both "$M \subset S$" and "$M \subset P$" are true, M counts as a piece of positive evidence for "$S \subset P$" — the existence of M, with its given relations with S and P, confirms the proposed inheritance relation from S to P, to a certain extent. On the other hand, if "$M \subset S$" is true but "$M \subset P$" is not, M counts as a piece of negative evidence — the existence of M, with its given relations with S and P, refutes the proposed inheritance relation from S to P, to a certain extent. As discussed previously, neither the confirmation nor the refutation is decisive in the sense that it cannot be revised in the future by other evidence.

Hempel's paradox does not appear in NARS, because a green shirt counts as neither positive evidence nor negative evidence for "Ravens are black", according to the previous definition. Just as our intuition tells us, in NARS the existence of a green shirt is *irrelevant* to whether ravens are black.

Now let us see how the amount of evidence is measured. Such a measurement, *weight of evidence*, is suggested by (Keynes, 1921). Intuitively, when we get new (relevant) evidence for a statement, the weight of evidence about that statement *increases*, because now our judgment are based on more evidence. From the definition of evidence given previously, we know when a term M becomes positive/negative evidence for statement "$S \subset P$". Ideally, if all available positive/negative evidence of the statement can be represented as two sets of such terms, respectively, it is natural to define the weight of positive and negative evidence as the *size* of the sets, respectively. Let us refer to them as w^+ and w^-, and call their sum, $w = w^+ + w^-$, as the weight of avaliable evidence.

Therefore, if we assign $w^+ = 5$ and $w^- = 3$ for "$S \subset P$", it means that the stated inheritance relation from S to P has been confirmed five times and refused three times, according to the system's experience. Of course, the system's actual experience is much more complex than the above ideal situation. Usually evidence is not completely confirmative or rejective, and pieces of evidence are not equally weighted. Furthermore, many statements are supported *indirectly* by statements derived from experience. These facts prevent us from using the above method to actually determine concrete truth values, but we can still use the method to define and interpret truth value abstractly. As a result, $w^+ = 5$ and $w^- = 3$ means, more accurately, that the system's belief on "$S \subset P$" is *as strong as* the relation has been confirmed five times and refused three times, even though the value is not actually determined in this way.

Though in principle all the information that we want to put into a truth value is representable in the $\{w^+, w\}$ (or $\{w^+, w^-\}$) pair, it is not always natural or convenient for many purposes. Instead of using absolute measurements, we often prefer relative measurements, such as real numbers in the interval [0, 1]. Fortunately, it is easy to define relative measurements in terms of the weight of evidence defined above.

Let us define the *frequency* of a statement, f, as w^+/w. Because w is the number of times that the proposed inheritance relation is checked, and w^+ is the number of times that the relation is confirmed, f indicates the "success frequency" of the inheritances relation between the two terms, according to the experience of the system.

Obviously, this measurement is often used in everyday life. It is also closely related to probability, though it is still different from probability under the traditional interpretations (Kyburg, 1970) — logical (degree of confirmation), empirical (relative frequency), and subjective (degree of belief). As mentioned before, in NARS truth value indicates the support the statement gets from evidence. Given the statement and the evidence, the value of f is uniquely determined. This is similar to the logical interpretation of probability suggested by Keynes (Keynes, 1921) and Carnap (Carnap, 1950). Different from them, in NARS the evidence is not explicitly expressed in a judgment, so f cannot be determined by logical analysis within the language. Instead, f is defined as the frequency of favorite evidence, which makes it similar to the probability under an empirical interpretation (Reichenbach, 1949). However, f is not the limit of the frequency, but its value at a certain moment in a certain system, thus it is subjective and context-dependent. These features are stressed by the subjectivists, though they refuse to explicitly ground probability on the frequency of favorable evidence (Savage, 1954).

To represent a truth value by a frequency value alone is not enough for NARS: in addition, the system needs to know the the value of w in order to figure out how to revise frequency with new evidence (Wang, 1993). Can we find a natural way to represent the necessary information in the form of a relative measurements, or, more specifically, as a *ratio*? Later we will see why we do not want to use w directly (though this is possible), but prefer a measurement in the $[0, 1]$ interval.

One attractive idea would be to define a "second-order probability". The frequency defined above can be considered to be an estimate of the "first-order probability" (of the given inheritance relation), and the second-order probability is used to describe how good the first-order estimate is (Paaß, 1991). However, under the assumption of insufficient knowledge, it makes little sense to talk about the "probability" that "the frequency is an accurate estimate of an objective first-order probability of the inheritance relation". Because NARS is always open to new evidence, it is simply impossible to decide whether the frequency of a judgment will converge to a point in the infinite future, not to mention where the point will be.

However, it makes perfect sense to talk about the "near future". What the system needs to know, from the value of w, is how *sensitive* a frequency is to new evidence — then the system can use this information to make a choice among competing judgments. If we limit our attention to a future of "fixed horizon", we can represent the information in w in a *ratio* form.

Let us consider what will happen at the arrival of a piece of new evidence, with a constant weight k. We define the system's *confidence*, c, on a judgment as $w/(w+k)$. For our current purpose, k can be any positive number. Intuitively, confidence is the ratio of the weight of the "all current evidence" to the weight of the "all evidence in the near future". It indicates how much the system knows about the inheritance relation — 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. The higher the confidence is, the harder it will be for the frequency to be changed by new evidence, but this does not mean that the judgment is "truer", or the "more accurate", because in an open system like NARS, the concept of a real or objective probability does not exist.

It is easy to calculate w and w^+ from f and c, and therefore the truth value of a judgment can also be represented as a pair of ratios $<f, c>$ (Wang, 1994a). In particular, "$S \subset P <1, 1>$" means that "$S \subset P$" is absolutely true. Though such a truth value cannot be achieved by finite amount of evidence, it serves as a limit and an idealized situation.

It is important to notice that f and c are two independent measurements, in the sense that given the value of either of them, the value of the other cannot be determined, or even bounded. Keynes argued for a similar relation between probability and weight of evidence (Keynes, 1921). Roughly speaking, frequency and probability indicate the relative balance between positive and negative evidence, which influences the system's preference among alternative conclusions; confidence and weight of evidence indicate the absolute amount of available evidence, which influence the system's sensitivity to new evidence.

Some authors claim that when "probability" is interpreted as "degree of belief" of an individual, and probability theory is used as a *normative* theory for how the individual should behave to maintain a consistent belief space, a probability distribution on the belief space is capable of representing the sensitivity mentioned above, because its effect eventually appears in the individual's preference among possible options in making a decision. Therefore, it can be captured by the range of belief changes on receipt of further evidence. Because belief changes can be properly handled by Bayes's theorem (and its variations, such as Jeffrey's rule), a new measurement is unnecessary (Cheeseman, 1985; Pearl, 1988; Spiegelhalter, 1986).

The problem in this argument, as shown in (Wang, 1993), is the assumption that all reevaluation of the probability distribution $P(x)$, caused by new knowledge E, can be put into the form $P(x|E)$, that is, by conditionalization on E. This assumption is not always valid, because in the above formula E must satisfy the following constraints: (1) E is a binary proposition, (2) E is already in the proposition space upon which $P(x)$ is defined, and (3) $P(E) > 0$.

Therefore, the susceptibility represented in the Bayesian approach, such as the *confidence* defined in (Pearl, 1988), only reflects the stability of a probability assignment to *certain relevant evidence*, and the restrictions upon new knowledge severely limits the learning ability of the system. Especially, they make the system only open to certain types of new knowledge, therefore cannot be used in systems designed under the insufficient knowledge and resources assumption, as defined previously.

Generally speaking, as argued in (Wang, 1993), the amount of evidence cannot be derived from a first-order probability distribution defined according to the evidence. Therefore, we do need a second measurement which is especially for this quantity.

5 How to Generate Inductive Conclusions?

Let us suppose that we know "Tomato is a kind of plant" and "Tomato is a kind of vegetable". Now "tomato", as a common instance of "vegetable" and "plant", becomes positive evidence for inductive conclusion "Vegetable is a kind of plant".

Formally, the induction rule of NARS looks like this:

$$\frac{\begin{array}{ll} M \subset P & <f_1, c_1> \\ M \subset S & <f_2, c_2> \end{array}}{S \subset P \quad <f, c>}$$

According to the definition of evidence, the term M now is, to a certain extent, a piece of evidence for the conclusion, and the truth value of the conclusion is a function of the truth values of the two premises. Similarly, "$P \subset S$" can also be assign a truth value, which is omitted in the following discussions, because it can be obtained by exchanging the order of the premises.

For our current example, "tomato" is M, "plant" is P, and "vegetable" is S. To determining the truth value of "*vegetable* $\subset$ *plant*" from the common instance "tomato" of the two terms, let us at first consider the following special situations.

1. When $f_1 = c_1 = f_2 = c_2 = 1$, M is a piece of (idealized) positive evidence for the conclusion. According to the previous definitions, in this case we have $w^+ = w = 1$ for

the conclusion — that is, $f = 1,\ c = 1/(1+k)$. For the "Ravens are black" example, here M is a black raven.

2. When $f_1 = 0,\ c_1 = f_2 = c_2 = 1$, M is a piece of (idealized) negative evidence for the conclusion. According to the previous definitions, in this case we have $w^- = w = 1$ for the conclusion — that is, $f = 0,\ c = 1/(1+k)$. For the "Ravens are black" example, here M is a non-black raven.

3. When $f_2 = 0$, M is not an instance of S. In this case, no matter it is an instance of P or not, it provides no evidence for the conclusion, therefore $w = 0,\ c = 0$, and f is undefined. For the "Ravens are black" example, here M is not a raven (but a shirt, for example).

4. When c_1 or c_2 is 0, one of the premises gets no evidential support, so the conclusion get no evidential support, neither. That means $w = 0,\ c = 0$. For the "Ravens are black" example, here either whether M is a raven or whether M is black is completely unknown.

From these boundary conditions of the truth value function for induction, if we assume all the variables take boolean values (either 0 or 1), we get $f = f_1$ and $w = AND(f_2, c_2, c_1)$, here AND is the boolean product of the arguments.

To extend the AND operator from boolean variables to variables in the [0, 1] interval (i.e., with boolean values as boundary values), we can use the so-called *T-norm* (Bonissone and Decker, 1986; Dubois and Prade, 1982; Schweizer and Sklar, 1983). T-norm is a binary function defined on real numbers in [0, 1]. It is monotonic, commutative, associative and has a boundary condition satisfying the truth table of the logical operator AND. For the current purpose, we also want it to be continuous and strictly increasing, so that changes in any one argument will cause a change in the function value. The most simple function that satisfy the above requirements are multiplication (Schweizer and Sklar, 1983).

Consequently, for the inductive conclusion, we have

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

where k is the constant introduced previously.

To apply this formula to the "tomato" example, we can see that the truth values of the two premises play different roles in induction. The frequency of "*tomato* $\subset$ *plant*", f_1, estimates the frequency of the conclusion, since we are taking the property ("being *plant*") of the special term *tomato* as a property of the general term *vegetable*. On the other hand, f_2, c_1 and c_2 *conjunctively* determines to what extent *tomato* can be counted as a piece of relevant evidence of the conclusion. It is because that if f_2 or c_2 is 0, *tomato* is not an instance of *vegetable* (so it cannot serve as evidence); or, if c_1 is 0, the first premise provides no information about the relation between *tomato* and *plant*, thus the conclusion gets no support, neither. Only when f_2, c_2, and c_1 are all equal to 1, can *tomato* be count as a piece of evidence (for the given conclusion) with a weight of 1 (because now "*tomato* $\subset$ *vegetable* $< f_2, c_2 >$" become "*tomato* $\subset$ *vegetable*"), and whether the evidence is positive or negative is completely determined by f_1.

Therefore, when an inductive conclusion is actually generated, the system does not treat all evidence as equal. For example, typical vegetables (with high f_2 and c_2 values) contribute more to the conclusion. On the other hand, the truth value function is established according to the relationship between w^+, w, f, and c, defined in idealized situations (where all pieces evidence are equally weighted, and are either completely positive or completely negative). In this way, the idealization is a necessary step in the design process of the system, and it also help us to understand the system's behavior.

It needs to be stressed again that the truth value of the conclusion indicates the support provided by the evidence, rather than measures how many vegetables are plants in the real world. If $f_1 = f_2$, $c_1 = c_2$, the system will assign the same truth value to "vegetables are plants" and "plants are vegetables". This may look ridiculous to us, but both of them are equally valid, given the evidence provided by "tomato" — we judge the second conclusion as less true than the first one, because we take other evidence (provided by pine, daisy, and so on) into account. We will see how the system does similar things in the following section, but before considering other evidence, the system believes the two conclusions to the same extent.

Because in NARS the truth value indicates the relation between a statement and available evidence, induction is "ampliative" in the sense that its conclusions are more general than its premises, but it is also "summative" in the sense that the conclusions claim no more support than they actually get from the premises. Therefore the traditional distinction between these two types of induction does not apply here (Cohen, 1989; Popper, 1959).

The system uses inductive conclusions to predict future situations, but it does not mean that their truth values tell the system what the "state of affirms" is in the "objective world". A system behaves according to its beliefs, not because they guarantee success (such guarantees are impossible, as Hume argued), but because it has to rely on its experience to survive, even though the experience may be biased or outdated — this is what "adaptation" means.

Another feature that distinguishes the above induction rule from other induction systems is that the rule is able to generate and evaluate an inductive conclusion at the same time.

Traditionally, the generating and evaluating of inductive conclusions (or hypotheses) are treated as two separated processes. The most well-known arguments on this issue are provided by Carnap and Popper, though their general opinions on induction are opposite (Carnap, 1950; Popper, 1959). The consensus is that from given evidence, there is no effective procedure to generate all the hypotheses supported by the evidence, therefore the discovery of a hypothesis is a *psychological* process, which contains an "irrational element" or "creative intuition". On the contrary, the evaluation of a given hypothesis, according to given evidence, is a *logical* process.

The above opinion is in fact implicitly based on the specific language in which the inductive process is formalized. In probability theory, there is no way to get a unique hypothesis H from given evidence E for the purpose of induction, because for every proposition X in the proposition space, $P(X|E)$ can be calculated, at least in principle. In first-order predicate logic, there are usually many hypotheses H that implies the given evidence E, and also consistent with background knowledge B. In both cases, we can use some heuristics to pick up a inductive conclusion that has some desired properties (simplicity, for instance), but this kind of selections are not derived from the definition of induction rule (Mitchell, 1980; Haussler, 1988).

In term logic, the situation is different. Here premises of an inductive inference must be a pair of judgments that share a common subject, and the premises uniquely determine an inductive conclusion. (Of course, there is also a symmetric inductive conclusion if we exchange the order of the premises.) Therefore, in NARS we do not need an "irrational element" or domain-dependent heuristics, and the discovery of a hypothesis, in the current sense, also follows logic.

It does not mean, however, that when given the same evidence, everyone should get identical inductive conclusions. One factor that cause "individual difference" is the constant k for "near future". Given the same evidence, a system with a larger k will assign lower confidence to inductive conclusions, because it considers what may happen in a further horizon. Consequently, such a system is more prudent, compared with a system with a smaller k (Wang, 1995c). For our current purposes, there is no best k for a implementation of NARS — it is a "personal parameter", and different values generate different behaviors. In the following discussions, let us take $k = 1$ for simplicity. Therefore, by "near future" we mean that "when the coming evidence has a unit weight".

6 How to Revise Inductive Conclusions?

From the above description, we see that NARS generates an inductive conclusion from a single piece of evidence, given by a pair of judgments. Inductive conclusions generated in this way have low confidence — $c \leq 1/2$ (when $k = 1$), according to the truth value function given above.

To increase the confidence of the conclusion, evidence from different evidence need to be accumulated. For example, let us assume that M_1 and M_2 are different evidence for "$S \subset P$", and they assign truth values "$\{w_1^+, w_1\}$" and "$\{w_2^+, w_2\}$" (in terms of weight of evidence) to the statement, respectively. If both M_1 and M_2 are taken into account, the truth value of the inductive conclusion should be "$\{w_1^+ + w_2^+, w_1 + w_2\}$", because weight of evidence is additive. Rephrasing the function in terms of frequency and confidence, we get the following *revision rule* for NARS:

$$\frac{\begin{array}{ll} S \subset P & <f_1, c_1> \\ S \subset P & <f_2, c_2> \end{array}}{S \subset P \quad <f, c>}$$

where

$$\begin{aligned} f &= \frac{c_1(1-c_2)f_1 + c_2(1-c_1)f_2}{c_1(1-c_2)+c_2(1-c_1)} \\ c &= \frac{c_1(1-c_2)+c_2(1-c_1)}{c_1(1-c_2)+c_2(1-c_1)+(1-c_1)(1-c_2)} \end{aligned}$$

This rule is applicable only when the two premises are based on different evidence. In NARS, a serial number system is used to approximately record the evidence that supports each judgment. See (Wang, 1995b) for details.

In the Bayesian approach, the evidence is accumulated by repeatedly applying Bayes's theorem to new evidence. Positive (negative) evidence is that which increase (decrease) the probability of the hypothesis in this process, and irrelevant evidence leave the probability unchanged. As discussed previously, this approach limits the evidence that is acceptable by the system. Besides, it also cause a paradox revealed by Popper (Popper, 1959). Let H be a hypothesis whose prior probability (according to the background knowledge of the system) is $P(H)$. If the system then gets a piece of positive evidence E_1 for H, the belief on it should be revised, according to Bayesian theorem, to become $P(H|E_1)$, which is higher that $P(H)$. After that, the system gets a piece of positive evidence E_2, which decreases the belief to $P(H|E_1 \wedge E_2)$. If $P(H|E_1 \wedge E_2)$ happen to be equal to $P(H)$, by definition, $E_1 \wedge E_2$ is irrelevant to H — altogether, it does not change the system's belief on H. However, intuitively we feel that the system knows more about H after learning both E_1 and E_2. What is wrong here?

In NARS, we can see that in this process f is first increased, then decreased, but c is increased by both E_1 and E_2. Therefore, the final result is more confident (because it is based on more evidence), though the influences of the two pieces of evidence on frequency cancel each other. In the Bayesian approach, both the f factor and the c factor are combined into a single probability distribution. It works fine for many purposes, but cannot handle revision properly, where the roles played by the two factors are different (Wang, 1993).

Let us continue to discuss the previous "vegetable–plant" example. Suppose after getting the "tomato" evidence (which equally support "Vegetables are plants" and "Plants are vegetables"), the system is told that "Pines are plants" and that "Pines are not vegetables". According to the induction rule given above, "pine" provides a piece of negative evidence for "Plants are vegetables", but is irrelevant to "Vegetables are plants". When these two inductive conclusions meet the corresponding conclusions generated from "tomato", "Plants are vegetables" obtains a frequency of 0.5 (because the positive and negative evidence have equal weight), and "Vegetables are plants" has a frequency of 1 (because all known evidence is positive). On the other hand, the former conclusion has a higher confidence than the latter — it is based on more evidence than the latter.

In this way, the induction rule generates a pair of inductive conclusions from two judgments that share subject, and the revision rule merges corresponding inductive conclusions to get more confident results. The revision rule can also be seen as a rule that resolve conflicts among beliefs.

Since the premises and conclusions of all the rules in NARS have the same syntax (as judgments defined previously), it is natural for the system to integrate different types of inference. Besides the induction rule and revision rule, NARS also has rules for deduction, abduction, exemplification, comparison and analogy (Wang, 1994a; Wang, 1995b), which are beyond the scope of this paper. Here we only need to mention that the premises used by the induction rule may be generated by the deduction (or abduction, and so on) rule, and that the conclusions of the induction rule may be used as premises by the other rules. In particular, the revision rule may merge an inductive conclusion with a deductive (or abductive, and so on) conclusion. Consequently, though NARS has an induction rule, it is not an "inductive logic", in the sense that it solves problems by induction only. The answers reported to the user are usually cooperative results of several rules of a multi-step inference. In NARS, all these rules are established according to the same semantics introduced previously, where the truth value of a judgment indicates the evidential support the judgment obtained. Different rules correspond to different ways to collect evidence for various inheritance relations (Wang, 1995b).

Though, as discussed previously, in NARS induction is not ampliative in a certain sense, the traditional distinction between "truth-preserving" and "ampliative" inferences is still there. In NARS, the confidence of deductive conclusions have a upper bound of 1, and we already know that the upper bound for induction is $1/(1+k)$, which is smaller than 1. If all premises are absolutely certain, so are their deductive conclusions, but no are their inductive conclusions.

Compared with other multi-strategy inference models using first-order predicate language (Michalski, 1993), attribute-value language (Giraud-Carrier and Martinez, 1995), or integrated symbolic/connectionist representation (Sun, 1995), the term logic model, proposed by Peirce and extended in NARS, puts different types of inference in the same framework in a more natural, elegant, and consistent manner.

From the above discussion, we see that conclusions in NARS are based on different amounts of evidence, and, generally speaking, conclusions based on more evidence are preferred, because their relative stability. However, since NARS is designed to be an open system, future evidence is always possible, therefore there is no way for the system to get "complete evidence" for an inductive conclusion.

A reasonable retreat is to use all evidence known to the system — the so-called "total evidence" (Carnap, 1950). Unfortunately, this is also impossible, because NARS has insufficient resource. The system has to answer questions under a time pressure, which makes exhaustive search in knowledge space not affordable.

Moreover, in NARS the time pressure is variable, depending to the request of the user and the existence of other information-processing tasks (Wang, 1995b). In this situation, even a predetermined "satisfying threshold" become inapplicable — such a threshold is sometimes too low and sometimes too high.

The control mechanism used in NARS is similar to "anytime algorithm" (Dean and Boddy, 1988). If the system is asked to evaluate the truth value of a statement, it reports the best conclusion (i.e., with the highest confidence) as soon as such a conclusion is found, then continue to look for a better one, until no resource is available for this task (see (Wang, 1995b) for how the system's resources are allocated among tasks). In this way, from the user's point of view, the system may change its mind from time to time, when new evidence is taken into consideration. The system will never say that "This is the final conclusion and I will stop working on the problem."

The above discussion is directly related to the "acceptance" problem in inductive logic (Kyburg, 1994). As put by Cohen, "what level of support for a proposition, in the light of available evidence, justifies belief in its truth or acceptance of it as being true?" (Cohen, 1989). In NARS, there is no such a thing as "accepted as being true". Judgments are true to different extent, and the system always follows the best-supported conclusion (compared with its rivals), no matter what its truth value is — the standard is relative and dynamic, not absolute and static. In this way, an inductive conclusion also benefits from the refutation of competing conclusions, which is stressed by the Baconian tradition of induction (Cohen,

1989) — though its truth value may not change in this process, its relative ranking becomes higher.

According to the definition given be Peirce, the difference among deduction, abduction, and induction is the position of the shared term in the two premises. This property of term logic makes it possible for NARS to combine different types of inference in a "knowledge-driven" manner. In each inference step, the system does not decide what rule to use, then look for corresponding knowledge. Instead, it picks up two pieces of most accessible knowledge (provided by its memory-management mechanism, see (Wang, 1995b)) which share a term, and decide what rule to apply according to the position of the shared term. In general, an inference process in NARS consists many steps. Each step carries out a certain type of inference, such as deduction abduction, induction, and so on. These steps are linked together in run-time in a context-dependent manner, so the process does not follow a predetermined algorithm.

Therefore, NARS is not an "inductive machine" which uses an effective algorithm to generate inductive conclusions from given evidence. Carnap's argument against the possibility of this kind of machine (Carnap, 1950) is still valid. However, this argument does not prevent us from building a computer system that can do induction. The system does not have a general purpose induction algorithm, but can solve problems under its knowledge and resource constraints, and in the problem-solving activities there are inductive steps.

7 Conclusion

In this paper we introduces the components of the NARS project that are related to inductive inference. This treatment of induction is characterized by the following features:

1. To use a term-oriented language, rather than the mainstream propositional/predicate language, to represent knowledge.

2. To interpret truth value as a measurement of evidential support, and to use two numbers (frequency and confidence) to represent it.

3. To define induction (as well as deduction and abduction) as inheritance-based inferences, in the form of an extended syllogism.

4. To calculate the truth value of inference conclusions (in induction, revision, and so on) according to the above definition of truth value.

5. To mix induction and other types of inferences in run time in a context dependent manner, and to treat inference processes as anytime algorithms so as to answer questions under a variable time pressure.

Though the components discussed previously are relatively simple when compared with other approaches of induction, they do naturally and consistently address many problems arising in the history of the study of induction.

Such a treatment of induction is not necessarily suitable for all situations — as discussed in the beginning of the paper, "induction" has different interpretations. This approach, or the NARS project as a whole, is specially designed for the situation where an adaptive system has to work under insufficient knowledge and resources. Such situation has special importance for artificial intelligence and cognitive science, for both theoretical and practical reasons (Wang, 1994b; Wang, 1995b).

Acknowledgment

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

References

Aristotle (1989). *Prior Analytics.* Hackett Publishing Company, Indianapolis, Indiana. Translated by R. Smith.

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

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

Cheeseman, P. (1985). In defense of probability. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, pages 1002–1009.

Cohen, L. (1989). *The Philosophy of Induction and Probability.* Clarendon Press, Oxford.

Dean, T. and Boddy, M. (1988). An analysis of time-dependent planning. In *Proceedings of AAAI-88*, pages 49–54.

Dubois, D. and Prade, H. (1982). A class of fuzzy measures based on triangular norms. *International Journal of General Systems*, 8:43–61.

Giraud-Carrier, C. and Martinez, T. (1995). An integrated framework for learning and reasoning. *Journal of Artificial Intelligence Research*, 3:147–185.

Good, I. (1983). *Good Thinking: The Foundations of Probability and Its Applications.* University of Minnesota Press, Minneapolis.

Haussler, D. (1988). Quantifying inductive bias: AI learning algorithms and Valiant's learning framework. *Artificial Intelligence*, 36:177–221.

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

Holland, J., Holyoak, K., Nisbett, R., and Thagard, P. (1986). *Induction.* The MIT Press.

Hume, D. (1748). *An enquiry concerning human understanding.* London.

Keynes, J. (1921). *A Treatise on Probability.* Macmillan, London.

Korb, K. (1995). Inductive learning and defeasible infernce. *Journal of Experimental & Theoretical Artificial Intelligence*, 7:291–324.

Kuhn, T. (1970). *The Structure of Scientific Revolutions.* Chicago University Press.

Kyburg, H. (1970). *Probability and Inductive Logic.* Macmillan, London.

Kyburg, H. (1994). Believing on the basis of the evidence. *Computational Intelligence*, 10:3–20.

Michalski, R. (1983). A theory and methodology of inductive learning. *Artificial Intelligence*, 20:111–116.

Michalski, R. (1993). Inference theory of learning as a conceptual basis for multistrategy learning. *Machine Learning*, 11:111–151.

Mitchell, T. (1980). The need for biases in learning generalizations. In Shavlik, J. and Dietterich, T., editors, *Readings in Machine Learning.* Morgan Kaufmann, San Mateo, California. 1990. Originally published as a Rutgers Technical report.

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

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann Publishers, San Mateo, California.

Peirce, C. (1931). *Collected papers of Charles Sanders Peirce*, volume 2. Harvard University Press, Cambridge, Massachusetts.

Popper, K. (1959). *The logic of Scientific Discovery.* Basic Books, New York.

Quinlan, J. (1986). Induction of decision trees. *Machine Learning*, 1:81–106.

Reichenbach, H. (1949). *The Theory of Probability.* University of California Press, Berkeley, California. Translated by E. Hutten and M. Reichenbach.

Savage, L. (1954). *The Foundations of Statistics.* Wiley, New York.

Schweizer, B. and Sklar, A. (1983). *Probabilistic Metric Spaces.* North-Holland, Amsterdam.

Spiegelhalter, D. (1986). A statistical view of uncertainty in expert systems. In Gale, W., editor, *Artificial Intelligence and Statistics*, pages 17–56. Addison Wesley, Reading.

Sun, R. (1995). Robust reasoning: integrating rule-based and similarity-based reasoning. *Artificial Intelligence*, 75:241–295.

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

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

Wang, P. (1994b). On the working definition of intelligence. Technical Report 94, Center for Research on Concepts and Cognition, Indiana University, Bloomington, Indiana. Available via WWW at *http://www.cogsci.indiana.edu/farg/peiwang/papers.html*.

Wang, P. (1995a). Grounded on experience: Semantics for intelligence. Technical Report 96, Center for Research on Concepts and Cognition, Indiana University, Bloomington, Indiana. Available via WWW at *http://www.cogsci.indiana.edu/farg/peiwang/papers.html*.

Wang, P. (1995b). *Non-Axiomatic Reasoning System: Exploring the Essence of Intelligence*. PhD thesis, Indiana University.

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