Toward a Logic of Everyday Reasoning

Pei Wang

Abstract: Logic should return its focus to valid reasoning in real-world situations. Since classical logic only covers valid reasoning in a highly idealized situation, there is a demand for a new logic for everyday reasoning that is based on more realistic assumptions, while still keeps the general, formal, and normative nature of logic. NAL (Non-Axiomatic Logic) is built for this purpose, which is based on the assumption that the reasoner has insufficient knowledge and resources with respect to the reasoning tasks to be carried out. In this situation, the notion of validity has to be re-established, and the grammar rules and inference rules of the logic need to be designed accordingly. Consequently, NAL has features very different from classical logic and other non-classical logics, and it provides a coherent solution to many problems in logic, artificial intelligence, and cognitive science.

Keyword: non-classical logic, uncertainty, openness, relevance, validity

1 Logic and Everyday Reasoning

1.1 The historical changes of logic

In a broad sense, the study of *logic* is concerned with the principles and forms of valid reasoning, inference, and argument in various situations.

The first dominating paradigm in logic is Aristotle's Syllogistic [Aristotle, 1882], now usually referred to as *traditional logic*. This study was carried by philosophers and logicians including Descartes, Locke, Leibniz, Kant, Boole, Peirce, Mill, and many others [Bocheński, 1970, Haack, 1978, Kneale and Kneale, 1962]. In this tradition, the focus of the study is to identify and to specify the forms of valid reasoning

Pei Wang
Department of Computer and Information Sciences, Temple University, Philadelphia, USA
e-mail: `pei.wang@temple.edu`

in general, that is, the rules of logic should be applicable to all domains and situations where reasoning happens, as "laws of thought".

In *syntax*, traditional logic is a *term logic*, with the following features:

Categorical statements: Each statement contains a *subject term* and a *predicate term* (each representing a category), and they are linked by a *copula* (representing a relationship of generalization).

Syllogistic rules: Each inference rule takes two *premises* (both are categorical statements) sharing a common term, and produces a *conclusion* (also a categorical statement) between the other two terms.

In *semantics*, traditional logic has the following features:

Binary truth-value: Each statement is either *true* or *false*, exclusively, as demanded by the Law of Non-contradiction and the Law of Excluded Middle.

Correspondence theory of truth: "To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true." [Aristotle, 2005]

Validity as truth-preserving: An inference rule is *valid* if and only if it always derives true conclusions from true premises.

Consequently, this type of logic focuses on the most salient form of reasoning, *binary deduction*.

Even though this tradition had made great contributions, its limitations became unacceptable when a symbolic logic was needed in the study of the foundation of mathematics, and consequently Frege, Whitehead, and Russell established first-order predicate calculus [Frege, 1999, Whitehead and Russell, 1910], which is now usually referred to as *classical logic*. This logic is similar to traditional logic in the semantic features listed above, but has very different syntactic features:

Function–argument statements: Each proposition contains a function and a number of arguments,[1]

Truth-functional inference rules: Each rule has premises and a conclusion that are only related in truth-value, not necessarily in content.

Consequently, this logic is not a "term logic", but a "predicate logic". Furthermore, it is a "mathematical logic" – not only is its *form* mathematical, but also its *subject matter*, as it was designed primarily to provide a logical foundation for mathematics [Haack, 1978, Kneale and Kneale, 1962]. The most important form of reasoning in mathematics is theorem proving, and in this process theorems are derived from axioms and definitions, following well-defined inference rules. To serve this purpose, Frege took a very strong "anti-psychologism" position, and argued that logic should study "the laws of truth" rather than "the laws of thought". As a result, logicians have moved away from everyday reasoning process, and focused their attention on abstract formal systems, which often has little to do with actual thinking in the human mind.

---

[1] Now the "function" is usually called "predicate", though it should not be confused with the "predicate term" in term logic, since they are very different in major aspects.

Though the majority of post-Fregean logicians accepts the anti-psychologism position, and no longer builds logic systems according to the human reasoning process in any sense, they nevertheless sometimes apply the logic systems to situations outside mathematics, under the implicit assumption that mathematical knowledge is human knowledge in the “purest” form, so valid reasoning in mathematics sets up an ideal case for the other domains to approximate. It is intuitively acceptable to treat reliable knowledge as “axioms”, so as to reveal their implications as “theorems”. Accordingly, psychologists, linguists, and many other researchers have been trying to apply mathematical logic into their study.

For instance, when explaining human reasoning process, there are two competing schools in psychology: the “mental logic” school [Braine and O’Brien, 1998] and the “mental model” school [Johnson-Laird, 1983]. Though these two schools have very different opinions on how humans reason, both theories actually come from classical logic, and their difference is just that the former selectively adopts the inference rules of classical logic, while the latter selectively adopts the semantic theory of classical logic. In all the related discussions, one question is rarely asked: if classical logic was not designed as a normative model of human reasoning at all, why to use it to judge the validity, or to explain the mechanism, of human reasoning?

Actually, it is well documented that the reality of human reasoning systematically differs from the prescription of classical logic. One example is Wason’s Selection Task [Wason and Johnson-Laird, 1972]: when human beings are asked to check the truthfulness of a statement, they often go after confirming evidence, while “according to (classical) logic” only counter-evidence matters. This result is widely cited as evidence of the “illogical” or “irrational” nature of human thinking.

Similar cases can be found in other domains. Popper’s claim that a scientific theory can only be falsified, but never verified [Popper, 1959] is based on the asymmetry between falsification and verification of a universal proposition in predicate calculus, just like in Wason’s Selection Task. As soon as a scientific theory is taken to be such a proposition, the conclusion follows. In linguistics, Montague attempted to provide a semantic theory for natural languages using the semantic model developed for mathematical logic [Montague, 1970]. It has not been very successful in practical applications, though nevertheless is still widely taken as a cornerstone of formal semantics in linguistics.

Artificial Intelligence (AI) comes into this discussion with a different goal: instead of explaining the human mind, here the main task is to build “thinking machines” [Turing, 1950, Feigenbaum and Feldman, 1963]. There have been many debates on the objective and methodology of AI [Kirsh, 1991, Wang, 2008], and among them there is the “logicist AI” school [Hayes, 1977, McCarthy, 1988, Nilsson, 1991]. In a broad sense, this approach suggests to identify the “laws of thought” that govern human thinking, formulate them as a logic, and then implement the logic in a reasoning system, so as to make computers to “think like a human”. Since here the application domain is not restricted to mathematics, it has been clear from the very beginning that classical logic is not the proper tool for the job, and the difficulty is on formalizing *everyday*, *real-world*, and *commonsense* reasoning [McCarthy, 1989]. Even so, classical logic has been taken as a starting point in many AI proposals. For example, Hayes suggested to formalize "Naive Physics" [Hayes, 1979] in first-order logic, and for several decades the CYC project has been formalizing human "common sense knowledge" in various domains in a variant of first-order predicate calculus [Lenat, 1995].

In summary, in AI and cognitive science (CogSci), "logic" has returned to its original subject matter, that is, reasoning in all domains, or in everyday reasoning, though the most common tool used is still the logic developed for a special type of reasoning in a special domain, that is, theorem proving in mathematics.

1.2 Issues in everyday reasoning

Is there any fundamental difference between "the logic of theorem proving" and "the logic of everyday thinking"? After all, mathematical logic has achieved great success in mathematics, computer science, and many other fields [Halpern et al., 2001]. Why can't it provide a normative model for reasoning in general?

This is a much-discussed topic [Birnbaum, 1991, Haack, 1996, McDermott, 1987, Minsky, 1990, Stenning and van Lambalgen, 2008], and a number of issues have been raised to show the limitations of mathematical logic when applied to everyday reasoning. Though the issues are explored under different considerations, each of them more or less shows a difference between the demand of classical logic and the reality of human thinking.

Though the existence of such differences are widely acknowledged, there are diverse attitudes on how to interpret them, where the opinions can be roughly classified into three schools:

Radical: "The difference shows a fundamental limitation of *logic*, and provides evidence for the conclusion that the human mind does not follow any logic. The logical approaches toward AI is doomed to fail."

Liberal: "The difference shows a fundamental limitation of *classical logic*, though it is still possible to be resolved within the framework of logic, by certain *non-classical logic*. Logical AI is still feasible if a proper logic is used."

Conservative: "The difference shows no limitation of classical logic at all, because the problem is ill-defined, unjustified, or beyond the scope of logic. Logical AI is still feasible if classical logic is used properly."

In the following, several major issues are briefly described, together with the non-classical logic systems proposed, either in logic or in AI. The corresponding "radical" and "conservative" responses are not presented here, though they will be addressed later in the article.

Uncertainty: In everyday reasoning, a statement is usually neither absolutely true nor absolutely false, but somewhere in between. Furthermore, since an intelligent system should be able to compare different possibilities, a three-valued logic (where a statement can be "uncertain") is not enough, and some type of numerical measurement of the uncertainty is often needed. Solutions to this issue include various forms of *probabilistic logic* [Nilsson, 1986, Adams, 1998] and *fuzzy logic* [Zadeh, 1983].

Relevance: Classical logic suffers from the notorious "paradox of material implication" – the "implication" defined in the logic does not match the intuitive meaning of "if–then", and it leads to various "logically correct" but intuitively problematic inference, where the premises and conclusions are unconnected in their contents. This issue triggered the development of *relevance logic* [Anderson and Belnap, 1975]. The problem has additional significance in AI, because no system can afford the computational resources to generate all the "true-but-useless" conclusions.

Openness: In everyday reasoning, the system cannot evaluate the truth-value of statements according to a constant set of axioms, but has to open to new evidence, which may challenge the existing beliefs. To work in these situations, one may depend on some "default rules" to get tentative conclusions in the absence of counter-evidences, and to revise these conclusions when counter-evidences show up. This is what *non-monotonic logics* attempt to do [McCarthy, 1989, Reiter, 1987]. Furthermore, if some contradictions cannot be resolved instantly, the logic has to tolerant them and to be *paraconsistent* [Priest et al., 1989].

Amplification: Classical logic only covers *deduction*, but there are also *induction*, *abduction*, *analogy*, and other types of inference that play crucial roles in everyday reasoning. These types of inference are often called "ampliative", since their conclusions seem to include knowledge that are not in the premises, which make them useful when the system has to solve problems beyond its knowledge scope. However, since these types of inference are not "truth-preserving" in the traditional sense, their validity has been a controversial topic, and many different solutions have been proposed, including various forms of *inductive logic* [Kyburg, 1970, Flach and Kakas, 2000].

Though the above-mentioned non-classical logics differ greatly in their details, they share the methodological assumption that *the limitations of classical logic can be overcome in a divide-and-conquer manner*. Each non-classical logic typically addresses one of the limitations, by extending or revising classical logic in certain aspect, while keeping the other classical features [Haack, 1996, Gabbay, 2007, Russell and Norvig, 2010]. Such an approach is intuitively appealing, but it leaves the theoretical question unanswered: how should these logics be coordinated in real-world situations?

This question has special significance for the emerging field of Artificial General Intelligence (AGI), which distinguishes itself from the mainstream AI by stressing the general-purpose nature of intelligence [Goertzel and Pennachin, 2007, Wang and Goertzel, 2007]. As the general-purpose nature of logic is exactly what AGI demands, there are AGI approaches that are fully or partially based on logic [Bringsjord, 2008, Goertzel et al., 2008, Gust et al., 2009, Wang, 2006]. Even so, what type of logic is suitable for AGI is still a topic under debate.

It may seem that the AGI researchers should wait for the logicians to design proper logical models for their tasks, and then implement them into computer systems, rather than trying to build new logics by themselves. It is not the case, because

AGI researchers often approach the problem of reasoning from special perspectives and with special considerations, so they may be able to explore opportunities that have not been considered by logicians.

Given the nature of the field, an AGI system needs to handle all the above issues in reasoning (and more), and there are two overall strategies:

Integration: The system uses multiple techniques (and some of them are logics), and its overall function is achieved by the coordination of these techniques [Goertzel, 2009].

Unification: The system uses a single logic, though other techniques (and some of them are logics) can be used as optional tools to achieve the system's overall function [Wang, 2004b].

The following description will introduce a concrete example of the unification strategy, and later compare it with the integration strategy.

1.3 Different types of logic

If everyday reasoning indeed has a logic, it must be able to handle the issues listed previously that cannot be properly treated by classical logic. Before getting into the details, let us first analyze the working environment of *everyday reasoning* by comparing it with that of *theorem proving*.

Reasoning is a process in which new *knowledge* is derived from existing knowledge, and this process costs computational *resources*, mainly processing time and memory space. Though the above description uses computer terminology, similar things can be said for the human mind. Now let us compare these two types of reasoning with respect to their assumptions on knowledge and resources.

Theorem proving occurs in an *axiomatic system*, where the set of axioms is predetermined, and each axiom is assumed to be *true*. The task of reasoning is to reveal the logical implications of the axioms, the theorems. To "prove" a theorem means to find a reasoning process that starts at the axioms and ends at the theorem, and in each step of the process a reasoning rule is used, with some axiom(s) and/or proven theorem(s) as premise(s). Whether a theorem is proven has nothing to do with the resources the process takes, as far as they are finite.

For a logic to be used in theorem proving, it is necessary for it to be *sound*, in the sense that all of its inference rules are *valid* as defined by its semantics, where *validity* means truth-preserving, that is, only deriving true conclusions when the premises are true. A related property of the axiom–theorem set is *consistency*, that is, contradiction free. It is also highly desired for a logic to be *complete*, meaning that all the truths in the domain can be proved as theorems by the logic. When all these properties are satisfied, the set of truth and the set of axioms and theorems coincide, and the logic fully satisfies our need.

If a reasoning system has all these features, I call it a "pure-axiomatic reasoning system", and say that it is based on the assumption of "sufficient knowledge and resources", since all relevant knowledge needed for the system to do its job is all embedded in its axioms at the beginning, and the system can afford the time–space required for its theorem-proving activity. I will call the logic governing this reasoning system a "pure-axiomatic logic" or simply "axiomatic logic".[2]

Compared to theorem proving, everyday reasoning serves a fundamentally different function for cognitive systems. Reasoning is needed to provide guidance for us to deal with the current situation and to prepare for the future situations, according to what we already know. However, since our empirical knowledge comes from past experience, there is no logical guarantee that they will always correctly predict the future, as argued by Hume [Hume, 1748]. Even though some knowledge is stable enough to be taken as "truth", it and its logical implications are still far from enough to answer the challenges we must face, not to mention that there are time-restrictions on how soon the answers are needed, so not all possible alternatives can be considered.

This real-world restriction has been referred to as "AIKR", for "Assumption of Insufficient Knowledge and Resources" [Wang, 1995, Wang, 2011]. Concretely, this assumption consists of three parts:

Finite: At any moment, the system only has a constant amount of computational capacity (the number and speed of processors, the size of memory and storage, etc.).

Real time: Tasks may appear at any moment, and each has time requirements attached (as deadline, urgency level, etc.).

Open: A task may have novel content, such as new evidence that conflicts with the existing beliefs, or new problem that is beyond the system's current knowledge scope.

It is not hard to recognize that AIKR is a restriction under which everyday reasoning is carried out, and we also hope AGI systems to work in such situations. Classical logic cannot be applied in such an environment, even approximately, simply because many major factors here are not (and should not be) under consideration in theorem proving at all.

If a reasoning system has to work in such an environment, what kind of "logic" it has to follow? If we take the "liberal" position, and believe that reasoning activities in such a situation still can show certain rationality, we have to conclude that even though various non-classical logics have been moving in this direction, they have not moved far enough from classical logic, in that each of them accepts AIKR *partially*, rather than *completely*. For this reason, they can be referred to as "semi-axiomatic logics", while what we need for everyday reasoning is a "non-axiomatic logic" that is designed to completely accept AIKR.

One common objection to this analysis is to deny the possibility of such a "non-axiomatic" logic – some people will agree to call such a system "a bunch of heuristics" or "rule-based", but disagree to call it a "logic". This response brings us back

[2] Please note that "axiomatic logic" does not mean that all the inference rules of the logic are derived from some axioms. Axiomatization at the meta-level (among inference rules) is not the same as that at the object-level (among domain knowledge).

to the fundamental notions: (1) What is "logic"? — It is the study of valid inference; (2) What is a valid inference rule? — It must be truth-preserving; (3) What is truth?

Now we have reached the core of this discussion. Classic logic is based on a correspondence theory of truth [Haack, 1978]. When the logic is designed and analyzed, the truth-value of a statement is *defined* with respect to a *model*, using a model-theoretic semantics [Barwise and Etchemendy, 1989]. However, when the logic is used in a reasoning system, the truth-value of a statement is *decided* with respect to its relation with the axioms or premises with established or assumed truthfulness, according to a proof-theoretic semantics [Schroeder-Heister, 2006]. To apply such a logic to a concrete domain means that there is domain knowledge whose truthfulness can be trusted, and all required conclusions can be derived from such a reliable foundation.

In everyday reasoning, on the contrary, by definition no such a basis can be found. All knowledge, including the initially given premises, may be challenged by future evidence, and the system does not know enough to perfectly solve all problems. No matter what the system does, its conclusions will be fallible. If validity is understood as "producing infallible conclusions", then the system cannot have this property when dealing with all the problems it has to face.

The common answer is to say that the above task is beyond the scope of logic, and what humans do in this situation only have psychological, but no logical, explanation [Hume, 1748, Popper, 1959]. What is missed by this answer is the possibility of another type of logic, based on a different sense of validity and truth. After all, we have the intuitive feeling that even when it is impossible to get infallible conclusions, some inference rules still seem more "reasonable" than the alternatives, and there can be vague consensus among human beings on what conclusions were "reasonable" given the derivation context, though these conclusions later turned out to be wrong when compared with further information.

A new form of rationality, *relative rationality*, has been proposed [Wang, 2011], by which a "rational" solution is the best one the system can found *under the current knowledge–resource restriction*. This idea resembles Simon's "bounded rationality" and some other ideas [Simon, 1957, Good, 1983, Cherniak, 1986, Anderson, 1990, Russell and Wefald, 1991, Gigerenzer and Selten, 2002]. What makes this new approach different is that it is instantiated by a formal logic designed to completely accept AIKR, and the logic has been mostly implemented in a computer system [Wang, 1995, Wang, 2006, Wang, 2013].

2 An AIKR-based Logic

NAL (Non-Axiomatic Logic) is the logic part of NARS (Non-Axiomatic Reasoning System), an AGI project aimed at a thinking machine that is fully based on AIKR (Assumption of Insufficient Knowledge and Resources) [Wang, 2006]. Since the details of NAL has been described in many publications, especially [Wang, 2013], in this chapter it is not fully specified, but used as an example of a new type of logic.

Using NAL as a concrete case will help us to clarify the issues in the study of logic addressed previously.

2.1 Validity and semantics

A key feature of NAL is its "experience-grounded" semantics [Wang, 2005], which realizes the notion of relative rationality. According to this semantics, the *truth-value* of a statement measures the support the statement gets from the available evidence collected from the system's experience. Since evidence can be either *positive* (agreeing with the statement) or *negative* (dusagreeing with the statement), a *binary* truth-value will not be informative enough for the system to choose among competing statements. Instead, a numerical representation becomes necessary.

Under this definition of truth-value, the "validity" of an inference rule of NAL still means "truth preserving", that is, the truth-value of the conclusion generated by the rule should correctly measure the evidential supported provided by the premises (with their own truth-values), without considering the other knowledge the system has. Unlike in a correspondence theory of truth, such a truth-value is not determined according to the "state of affairs" in the world or a model.

Since in a valid inference step the premises must provide evidence for the conclusion, they must be *relevant in content*. Therefore NAL cannot use the traditional "truth-functional" inference rule, where if a proposition in the premise or conclusion of a valid step is replaced by another one with the same truth-value, the inference remains valid. In NAL it is no longer the case because the evidence supporting one statement may not support the other statement to the same extent merely because the two statements have the same truth-value.

Experience-grounded semantics is very different from model-theoretic semantics. It bears some similarity to proof-theoretic semantics [Schroeder-Heister, 2006] in spirit, though in NAL the reasoning process is no longer a "proof" that decides the truth-value of the conclusion *conclusively*. Instead, in NAL the truth-value of a conclusion is evaluated in each step *inconclusively*, since it can always be revised by further consideration with new evidence. Therefore in NAL "truth" is fundamentally *subjective* and *changeable*, though by no means *arbitrary*. Such a truth-value is in coherent with AIKR, since it only depends on available evidence that comes from the system's past experience, and is obtained using the resources allocated to the relevant reasoning tasks.

To be compatible with such a semantics, the formal language of NAL must allow the *evidence* of a statement to be naturally defined and measured.

NAL uses a formal language *Narsese* for internal representation and external communication. Narsese is an "term-oriented language", also known as "categorical language", as exemplified by Aristotle's logic [Aristotle, 1989]. Different from the "function–arguments" format of classical logic, a sentence in a term-oriented language has the format of "subject–copula–predicate" format, as mentioned previously. In the simplest situation, a term is just an internal identifier of a category or concept. In the following description, English common nouns are used as terms.[3]

The most basic copula of NAL is "inheritance", expressed by '$\rightarrow$'. In its idealized form, it is a binary relation between terms, and is defined by being reflexive and transitive. The intuitive meaning of inheritance statement "$S \rightarrow P$" is that "S is a specialization of P", and equivalently, "P is a generalization of S". In this way the intuitive meaning of Narsese statements like "$raven \rightarrow bird$" and "$water \rightarrow liquid$" can be understood.

From a given (finite) set of inheritance statements, called the system's "idealized experience", some other inheritance statements can be derived according to the transitivity of the copula. Formally, the transitive closure of the idealized experience forms the system's "knowledge", or "truths". For a given term T in the system's experience, the set of its known specializations (plus itself) is called its *extension*, T^E, and the set of its known generalizations (plus itself) is called its *intension*, T^I. For example, if "$water \rightarrow liquid$" is in the system's experience, then $water \in liquid^E$ and $liquid \in water^I$. It can be proved that "$S \rightarrow P$" if and only if $S^E \subseteq P^E$, as well as $P^I \subseteq S^I$.

Now we can move from binary statements to multi-valued statements by using the former to define the latter. For a statement "$S \rightarrow P$", its positive evidence consists of the terms in $E^+ = (S^E \cap P^E) \cup (P^I \cap S^I)$, because as far as these terms are concerned, the statement is correct; the negative evidence of the statement consists of the terms in $E^- = (S^E - P^E) \cup (P^I - S^I)$, because as far as these terms are concerned, the statement is incorrect.

The amount of positive, negative, and total evidence are defined as $w^+ = |E^+|$, $w^- = |E^-|$, and $w = w^+ + w^- = |S^E \cup P^I|$, respectively. The truth-value of the statement is represented by a pair of real numbers $\langle f, c \rangle$ in $[0,1] \times (0,1)$, where f is the *frequency*, defined as w^+/w, that is, the proportion of positive evidence among all evidence, and c is the *confidence*, defined as $w/(w+k)$, that is, the proportion of current evidence among all evidence after the coming of new evidence of amount k, where k is a constant parameter. In the following discussion, we take $k = 1$, that is, the current evidence is compared with a unit amount to indicate how much evidence the system already has on the statement.

Given this extension of truth-value, whether a term is in the extension or intension of another term is also a matter of degree. A statement with truth-value is called a *judgment*, and judgment "$S \rightarrow P \langle f, c \rangle$" indicates that S is in the extension of P, and P is in the intension of S, both to the extent measured by $\langle f, c \rangle$. The *meaning* of a term is determined by its extension and intension, i.e., the system's knowledge on its relations with other terms.

Now we can see why the semantics of NAL called "experience-grounded": given an idealized experience, the truth-value of the statements and the meaning of the terms are all determined accordingly. Since experience stretches in time, truth-value and meaning may change, and not necessarily converge, since no restriction is made on the content of the system's future experience.

[3] This usage does not suggest that such a term will have the same meaning as what the word means to an English speaker, but that their meanings have overlap to certain extent.

The way truth-value and meaning are *defined* is not how they are actually *obtained* when the logic is used. Under AIKR, the *actual experience* is a stream of Narsese judgment. The derived judgments are generated by the inference rules, each with a truth-value indicating the evidence provided by the premises. Here the function of the semantics is to interpret the input and output judgments, as well as to guide the design of the inference rules. No knowledge, given or derived, has the status of an "axiom", that is, with a known and invariable truth-value. This is why NARS and NAL are called "non-axiomatic".

2.2 Basic inference rules

As a term logic, an inference rule of NAL typically takes two premises that share a common term, and generates a conclusion between the other two terms. To be concrete, let us say that the first premise is between M and P with truth-value $\langle f_1, c_1\rangle$, the second premise is between M and S with truth-value $\langle f_2, c_2\rangle$, and the conclusion is "$S \to P\,\langle f, c\rangle$", where f and c are calculated from the truth-values of the premises by a truth-value function.

The truth-value function is designed by first treating all the involved quantities as Boolean variables that only take values in $\{0, 1\}$. Then, Boolean functions are established among these variables according to the semantics. Finally, the Boolean functions are extended into real-number functions using the *product triangular norm*:

$$not(x) = 1 - x,\;\; and(x,y) = x \times y,\;\; or(x,y) = not(and(not(x), not(y)))$$

The most straightforward rule is the *deduction* rule:

$$\{M \to P\,\langle f_1, c_1\rangle, S \to M\,\langle f_2, c_2\rangle\} \vdash S \to P\,\langle f, c\rangle$$

This rule extends the transitivity of the inheritance copula from the binary case to the general (multi-valued) case. Here the binary inheritance relation can be seen as a special case of the multi-valued version when f is 1 and c is converging to 1. So the truth-value function of this rule is given by

$$f = and(f_1, f_2),\; c = and(f_1, c_1, f_2, c_2)$$

Following the insight of Peirce [Peirce, 1931], the *induction* rule and the *abduction* rule are obtained by switching the conclusion of the *deduction* rule with each of its premises, respectively. After renaming the terms and variables, they are:

$$\text{Induction: } \{M \to P\,\langle f_1, c_1\rangle, M \to S\,\langle f_2, c_2\rangle\} \vdash S \to P\,\langle f, c\rangle$$

$$\text{Abduction: } \{P \to M\,\langle f_1, c_1\rangle, S \to M\,\langle f_2, c_2\rangle\} \vdash S \to P\,\langle f, c\rangle$$

Unlike deduction, these two rules are invalid in their binary form (i.e., when the truth-values are omitted and all the statements involved are taken to be "true"). However, in NAL they are valid, as they exactly correspond to the extensional component and the intensional component of evidence of the conclusion, respectively. According to the definition of evidence, for induction and abduction we have, respectively

$$\text{Induction: } w^+ = and(f_1, c_1, f_2, c_2),\ w^- = and(not(f_1), c_1, f_2, c_2)$$

$$\text{Abduction: } w^+ = and(f_1, c_1, f_2, c_2),\ w^- = and(f_1, c_1, not(f_2), c_2)$$

Induction and abduction are "weak inference", since in their conclusion $w < 1$, so $c < 0.5$ (when $k = 1$). On the other hand, deduction is "strong inference", since the confidence of its conclusion takes 1 as the upper bound. In this way, the traditional "deductive inference vs. non-deductive inference" distinction is still made in NAL, though it is quantitative, in that the conclusions of deduction are "stronger" (less sensitive to new evidence) than those of induction and abduction.

When the same statement is supported by disjoint bodies of evidence, there will be two truth-values for the same statement. Whenever such a pair of judgments is located, the *revision* rule of NAL generates a conclusion that is based on the pooled evidence. The following truth-value function comes from the additivity of amount of evidence:

$$w^+ = w_1^+ + w_2^+,\ w^- = w_1^- + w_2^-,\ w = w_1 + w_2$$

The *revision* rule is the only inference rule in NAL whose conclusion has a higher confidence value than those of the premises. Through this revision process, judgments become stronger by merging with each other, and evidence from different sources is accumulated.

With the above rules, NAL can be used to answer questions. For an "yes/no" question on statement "$S \to P$", the *choice* rule picks a matching judgment with the highest confidence value; for an "what" question of the form "$S \to ?$" or "$? \to P$", this rule picks a matching judgment with the highest *expectation* value e, where $e = c(f - 0.5) + 0.5$.

When a question cannot be directly answered by available judgments, the syllogistic rules can be used for *backward inference* to derive questions from the existing questions and relevant judgments, under the condition that the answers of the derive questions can contribute to answers of the original questions.

2.3 Layered structure of NAL

In the current design [Wang, 2013], NAL is introduced in 9 layers, NAL-1 to NAL-9. Each layer extends the grammar rules, semantics, and inference rules to increase the expressive and inferential power of the logic, while respecting AIKR.

NAL-1 has been mostly described above. It is the simplest non-axiomatic logic, where the language includes inheritance statements between atomic terms, the semantics is experience-grounded, and the main inference rules are *deduction, induction, abduction, revision*, and *choice*.

Restricted by the chapter length, the other layers are only briefly described in the following. For the details, see [Wang, 2013].

NAL-2 introduces a *similarity* copula, '$\leftrightarrow$', as a symmetric version of inheritance. In its binary form, $S \leftrightarrow P$ is defined as $(S \rightarrow P) \wedge (P \rightarrow S)$, and the evidence of either inheritance statement is taken as evidence of the similarity statement.

With two copulas, the syllogistic rules of the system have three new forms.

- The *comparison* rule is a weak rule like induction and abduction, except that its conclusion is a similarity statement, obtained by comparing the two terms with a third term in their extension or intension;
- The *analogy* rule uses a similarity statement to carry out term substitution in an inheritance statement;
- The *resemblance* rule extends the transitivity of the similarity copula from binary to multi-valued.

These two copulas form a conceptual hierarchy, with *inheritance* for the "vertical" relations, and *similarity* for the "horizontal" relations. NAL-2 also introduces two special types of term to indicate the "floor" and "ceiling" of this hierarchy, respectively. An *extensional set* $\{T\}$ cannot be further specialized, and an *intensional set* $[T]$ cannot be further generalized. For example, $\{Aristotle\}$ ("Aristotle-like") represents the concept whose extension is fully specified by a single instance, and $[black]$ ("black things") represents the concept whose intension is fully specified by a single property. The term in the set roughly correspond to a proper noun and an adjective in English, respectively.

NAL-3 introduces *compound terms*, each of which is formed by a connector and a few component terms. In particular, in this layer four set-theoretic compounds are defined, together with the inference rules that compose and decompose them.

The compound term $(T_1 \cap T_2)$ is the *extensional intersection* of terms T_1 and T_2, and a composition rule is

$$\{M \rightarrow T_1 \langle f_1, c_1 \rangle, M \rightarrow T_2 \langle f_2, c_2 \rangle\} \vdash M \rightarrow (T_1 \cap T_2) \langle f_1 f_2, c_1 c_2 \rangle$$

The compound term $(T_1 - T_2)$ is the *extensional difference* of terms T_1 and T_2, and a composition rule is

$$\{M \rightarrow T_1 \langle f_1, c_1 \rangle, M \rightarrow T_2 \langle f_2, c_2 \rangle\} \vdash M \rightarrow (T_1 - T_2) \langle f_1(1 - f_2), c_1 c_2 \rangle$$

These compositional rules are not exactly "syllogistic", because they do not build new relations among the given terms. However, since they still demand the premises to have a common term, they can be considered as syllogistic in a broader sense of the notion.

There are also *intensional intersection* and *intensional difference* that are defined symmetrically to the above two compounds, and in them the common term is in the intention of the other terms. Furthermore, *extensional set* and *intensional set* are extended to allow any number of components.

NAL-4 transforms various conceptual relations into the inheritance relation. For example, if there is an arbitrary relation R between A and B, it can be expressed in

Narsese as $(A \times B) \rightarrow R$, where the subject of the inheritance statement is a *product* of A and B. The same information can be equivalently expressed as $A \rightarrow (R \,/ \diamond B)$ and $B \rightarrow (R \,/\, A \diamond)$, where the predicate is an *extensional image* with an indicator '$\diamond$' for the position of the subject in the relation. There is also an *intensional image* that can equivalently represent $R \rightarrow (A \times B)$ as $(R \setminus \diamond B) \rightarrow A$ and $(R \setminus A \diamond) \rightarrow B$.

In this way, NAL can express and process arbitrary relations, while the inference rules are still only defined on the two copulas. The other conceptual relations are represented as terms, with experience-grounded meaning. On the contrary, copulas are not terms, and their meaning is fully specified by the inference rules, independent of the experience of the system.

NAL-5 allows a statement to be handled as a term. NAL treats verbs like "know" and "believe" as a relation between someone and a statement, and let the meaning of the relation be acquired from the system's experience, rather than to define their meaning within the logic, like in epistemic logic [Hendricks and Symons, 2015].

Two statement-level copulas are introduced at this level. *Implication* ($\Rightarrow$) means "can be derived from", and *equivalence* ($\Leftrightarrow$) means "can derive each other". Since they are isomorphic to the term-level copulas *inheritance* ($\rightarrow$) and *similarity* ($\leftrightarrow$), respectively, many inference rules in the lower levels can be mapped into this level. For example, the statement-level *deduction* rule has the form of

$$\{M \Rightarrow P \langle f_1, c_1 \rangle, S \Rightarrow M \langle f_2, c_2 \rangle\} \vdash S \Rightarrow P \langle f, c \rangle$$

and its truth-value function is the same as the deduction rule of NAL-1.

This level also introduces statement connectors *negation* ($\neg$), *conjunction* ($\wedge$), and *disjunction* ($\vee$). Though their intuitive meaning is the same as in propositional logic, in NAL they are not defined by truth tables. The truth-value of $(\neg S)$ is obtained by switching the positive and negative evidence of S. *Conjunction* and *disjunction* are defined as isomorphic to *extensional intersection* ($\cap$) and *intensional intersection* ($\cup$), respectively. The inference rules on them are defined accordingly.

NAL-6 introduces *variable terms* into NAL. A variable term does not identify a concept, but serves as a "symbol" of another term, so it may identify different concepts in different situations. For instance, statement "$(\$x \rightarrow P) \Rightarrow (\$x \rightarrow Q)$" expresses "Whatever in the extension of P is also in the extension of Q", where $\$x$ is an *independent variable* representing an arbitrary term in P^E. Similarly, statement "$(\#x \rightarrow P) \wedge (\#x \rightarrow Q)$" expresses "There is something in the extensions of both P and Q", where $\#x$ is a *dependent variable* representing an anonymous term in $P^E \cap Q^E$.

Many inference rules of NAL-5 can be extended to handle variable elimination, introduction, and unification in NAL-6 by adding a substitution step before or after the applying of the NAL-5 rule.

Using variable terms, NAL can carry out hypothetical inference on abstract concepts, then apply the result to different concrete situations by interpreting the variables differently. In particular, NAL can serve as the meta-logic of an arbitrary logic, by representing the axioms and theorems of the latter as terms, and the inference rules of the latter as implication statements. In this way, NARS can have axiomatic or semi-axiomatic subsystems, outside the restriction of AIKR.

NAL-7 directly supports temporal inference on *events*, which are statements with time-dependent truth-value. At this layer, two primitive temporal relations, *sequential* and *parallel*, are embedded into Narsese by combining with connectors (like *conjunction*) and copula (like *implication*). When the inference rules are given premises with temporal attributes, the temporal factor and the logical factor are handled independently, then the conclusion is determined by both results.

NARS also uses an internal clock (with its reasoning cycle as unit) to get a sense of "subjective time". The system not only can reason "about time", but also "in time", as the *present* will gradually become about the *past*, while the *future* will become the *present*. In this aspect, NAL is very different from conventional *temporal logics* [Vila, 1994], which assume the reasoning system itself is working outside the stream of time, by treating the tense of a proposition as its intrinsic attribute.

NAL-8 specifies *procedural inference*. Using the idea introduced by logic programming [Kowalski, 1979], certain terms can have a *procedural interpretation* by associating with executable programs. In NAL, an *operation* has the form of $op(a_1 \dots a_n)$, where op is the *operator* that is associated to a program, and $(a_1 \dots a_n)$ is a list of *arguments* that are the input and output of the program. When an operation is involved in reasoning, it is treated as statement $(a_1 \times \dots \times a_n \times SELF) \rightarrow op$, and treated like the other statements. Here *SELF* is a special term representing the system itself, as an operation logically is a relation among the system and the arguments.

Before this layer is added, NAL can handle two types of inference tasks: to absort a *judgments* and to answer a *questions*. At this layer, a third type is added: to achieve a *goal*. A goal is a statement that the system desires to realize. Each event E has a *desire-value* attached, which is defined as the truth-value of $E \Rightarrow D$, where D is a *virtual* term representing the desired situation. In this way, the desire-values are conceptually transformed into truth-values, and are handled accordingly.

When the desire-value of an event is high enough, the system may make the decision to turn it into a goal to be actively pursued. A goal may generate derived goals via backward inference, and this process can repeat recursively until each derived goal is either an executable operation, or already satisfied (as far as the system knows). After the actual execution of these operations, the system revises its related knowledge according to the feedback or observation to reflect the changes.

NAL-9 enables the system to perceive and control itself via a set of *mental operations* that can sense and act on the system itself under the control of the inference process. Since this layer makes no change in the grammar rules, semantics, and inference rules, it can be considered either as an extension of the logic or as part of the implementation of the logic, so it will not be further described here.

2.4 Computer implementation

As the logic followed by NARS, NAL specifies what can be expressed in the system and what can be derived by the system. As a computer system, NARS has the following major parts:

- an input/output interface that realizes the grammar rules of Narsese,
- an inference engine that realizes the inference rule of NAL,
- a memory structure that contains the judgments, questions, and goals,
- a control strategy that selects premises and rules in each inference step.

All the above components are designed under AIKR, implemented in an open source project[4], and discussed in various publications [Wang, 2006, Wang, 2013, Hammer et al., 2016]. Since this chapter focuses on NAL, the above parts of NARS will not be discussed here.

3 Discussion and Comparison

The previous section provides a summary of NAL, as defined in [Wang, 2013]. With this preparation, now we can analyze several fundamental issues in logic.

3.1 Logic and valid reasoning

When challenging the Aristotelian tradition of logic, the "mathematical logic" paradigm changed the study of logic fundamentally. Though the fruitfulness of this change is undoubtable, it also has the undesired effect of leading the study away from its original and general objective, that is, to provide normative models for reasoning processes in various situations. Instead, most of the works have been focused on the reasoning in mathematics, especially in theorem proving. Even in the field of "philosophical logic", which covers non-mathematical reasoning, mathematical logic is still widely taken as the norm [Grayling, 2001]. Consequently, the reality of human reasoning has been largely ignored, with the justification of anti-psychologism.

Many researchers have expressed their disapproval of this situation, and argued for a closer relationship between logic and human thinking. In recent years, representative opinions can be found in [Gabbay and Woods, 2001, Gaifman, 2004, Hanna, 2006, Stenning and van Lambalgen, 2008]. This is also the driving force for the various non-classical logics to move away from classical logic.

NAL has been driven by a similar motivation, though it has gone much further than the previous non-classical logics. Instead of remedying one limitation of clas-

[4] At `https://github.com/opennars/opennars/`.

sical logic, NAL challenges the universality of the traditional definition of *validity* in reasoning. The reason to redefine a well-accepted notion is that what is "valid" should be judged according to the role reasoning plays for the reasoner.

As discussed previously, theorem proving serves the role of revealing the hidden implications of axioms in a constant and closed system, so "valid reasoning" means "truth-preserving" in the sense that whenever the axioms are applicable in a situation, so do the theorems. Therefore the traditional notion of validity is indeed the right choice for mathematical logic.

On the other hand, "everyday reasoning" (or call it "commonsense reasoning", "evidential reasoning", "real-world reasoning", "empirical reasoning", and so on) happens as an adaptive system's attempt to predict the future according to the past. Since in the real world the future will never be exactly the same as the past, this prediction cannot be faultless. Therefore the traditional notion of validity is inapplicable in such a process. However, it does not mean that there cannot be normative model here. Not all changes in a system qualify to be considered as "adaptive". As the selective result of the evolution process, the regularity observed in human thinking must represent certain form of optimality. In a *relatively stable* environment (which changes slower than the system's behaviors), the best strategy is to behave *as if* the current situation is like the past, even though it is known that this strategy will not always make correct productions.

Except in axiomatic systems, our judgments about truth are all *evidence-based*. Even in cases where we feel that we are directly checking a statement again *the reality*, we are actually checking the statement again *a description of the reality*, which not only depends on the reality as it is, but also depends on our sensory organs and our interpretation of the sensory input, which in turn depends on our conceptual repository, motivation, attention, and other factors. It is for this reason that in NAL "truth-value" is defined as "evidential support". Accordingly, valid inference rules are "truth-preserving" in the sense that the conclusion is based on the evidence provided by the premises.

In this sense, the non-deductive rules of NAL are truth-preserving, rather than "ampliative". Taking enumerative induction as an example. From "Tweety is a bird" and "Tweety flies" to derive "All birds (in the universe) fly" is ampliative (as the conclusion says more than the premises), but to derive "There is a bird that flies (as far as I know)" is not. The latter can be represented as "Bird flies $\langle t \rangle$" where t numerically represents something like "confirmed once", so that the conclusion does not say more than what is provided by the premises, as far as the truth-value "$\langle t \rangle$" is attached to the conclusion. In NARS, the really "ampliative" step is not this type of inference, but the system's usage of this conclusion when answering the question "Will the next bird I meet be able to fly?" – what is hoped is a sure answer, but the system can only provide its belief, which has limited evidential support and is fallible. However, it is exactly this response that let the system open to novel problems for which no sure solution can be found. It is a property of truly intelligent systems, which is not possessed by conventional computer systems.

What is called "everyday reasoning" in this chapter is similar to what is traditionally labeled as "inductive reasoning", which has been widely characterized as "ampliative", so is "falsifiable" and therefore "invalid". The practice of NAL shows that under proper treatment, this type of reasoning (which is not limited to induction) can be justified according to a different notion of validity. Compared to theorem proving, which has been properly formalized in mathematical logic, the type of logic exemplified by NAL is closer to the need of AI and CogSci. Though fundamentally different from classical logic, NAL qualifies to be called a "logic", just like non-Euclidean geometry is fundamentally different from Euclidean geometry, but still qualified to be called a "geometry".

3.2 Defining evidence

To establish a logic for *evidential* reasoning, the definition of "evidence" plays a crucial role.

This topic has been studied for many year in various domains. What observations will increase or decrease our belief on a general statement like "Ravens are black"? In first-order predicate calculus, this statement is formalized as "$(\forall x)(Raven(x) \supset Black(x))$" and an intuitive criterion, "Nicod's Criterion", is to take an object c as confirming (positive) evidence if $Raven(c) \wedge Black(c)$ is true (i.e., a black raven), and as opposing (negative) evidence if $Raven(c) \wedge \neg Black(c)$ is true (i.e., a non-black raven). If $Raven(c)$ is false, c (a non-raven) provides no evidence for the statement. However, as revealed by Hempel [Hempel, 1943], "$(\forall x)(Raven(x) \supset Black(x))$" is logically equivalent to "$(\forall x)(\neg Black(x) \supset \neg Raven(x))$" so, the two statements should have the same evidence. As according to Nicod's Criterion a non-black non-raven (such as a red apple or a white sock) is confirming evidence of the latter, it should also confirm the former. This result challenges Nicod's Criterion, as well as our intuition.

As analyzed in [Wang, 2009], according to Nicod's Criterion "Ravens are black" and "Non-black things are not raven" have the same *negative* evidence (*non-black ravens*), but different positive evidence (*black ravens* and *non-black non-ravens*, respectively). Since in NAL a truth-value depends on *both* positive and negative evidence, the two statements do not have the same truth-value, so are not equivalent. They are indeed equivalent in classical logic, because there "true" means "no negative evidence", and positive evidence does not matter.

The same difference between everyday reasoning and mathematical reasoning can be found in the well known psychological experiment "Wason's Selection Task" [Wason and Johnson-Laird, 1972]. When required to select cases to evaluate the truthfulness of a given general statement, people tends to select potentially confirming cases, while "according to logic" they should only select potentially disconfirming cases.

In the discussion about how scientific statements are evaluated, Popper argued that since they are universally quantified propositions, they cannot be verified by confirming examples, but can only be "falsified" by counter-examples. He went further to propose "hypothetical-deduction" as a general methodology of science, and interpreting "conforming" as "eliminating potential falsification".

Though the above debates happen in logic, psychology, and philosophy, respectively, they can all be reduced to the same question: when the truthfulness of a general statement is evaluated by checking concrete cases, should we only consider negative cases, or also positive ones?

According to the previous analysis, my conclusion given to all these debates is simple and clear: there are two types of domains using two different types of logic. In mathematics, usually only negative evidence matters. A mathematical *hypothesis* can be turned into a *theorem* only by proofs, not by confirming cases, no matter how many of them have been found. On the other hand, a *hypothesis* in an empirical science or everyday life can gradually evolved into a "law" with the accumulation of confirming evidence, as formalized in NAL.

To define truth-value as evidential support, another requirement for the inference rules is that the premises and the conclusion to be related in *content*, rather than merely in truth-value. While most relevant logics attempt to achieve this by revising predicate logic, NAL solves this issue using a more elegant solution provided by term logic, that is, by formalizing all inference rules as extended syllogism, where the the premises and the conclusion have shared terms. Though the similarity of Aristotle's syllogistic and modern relevance logic has been noticed by many logicians [Steinkrüger, 2015], few people has suggested that modern relevance logics should be designed as a term logic.

To replace the predicate logic framework by the term logic framework is not merely a change in notations, but a change in opinion on the nature of reasoning or inference. In propositional calculus and predicate calculus, an inference process is a chain of steps, in each of which the truth-value of a conclusion is decided purely according to its truth-functional relation with the premises. In a term logic, on the other hand, inference is primarily based on *the transitivity of the copula*, from which the truth-value relation between premises and conclusions is derived. In such a logic, the semantic relevance among premises and conclusions is a necessary consequence, not a property that needs to be added in by special mechanisms.

The basic form of copula in NAL, *inheritance*, plays no special role in predicate logic. Though it can be defined as a special predicate name, the basic inference rules are not based on it, but on the truth-value relations among propositions. This arrangement is allowed, even preferred, in mathematics, but unsuitable in everyday reasoning. This explains why in AI many logic-based systems make the additional arrangement to give the "is-a" relation a special treatment, such as in description logic [Baader et al., 2003], though few of them completely bases its reasoning on this relation and its variants, as in NAL.

Since induction, abduction, and analogy are invalid in theorem proving, but can be interpreted as evidential inference and conceptual substitution, they are more naturally formalized in term logic. However, because of the dominance of predicate logic in the field, in recent years they are usually formalized in the framework of predicate logic [Flach and Kakas, 2000]. This treatment directly leads to the consequence that they can only be used for *hypothesis generation*, but not *hypothesis* *evaluation*, mainly because the notion of *evidence* cannot be properly defined in predicate logic, where all logical relations are specified as purely truth-functional.

3.3 Various types of uncertainty

To handle various types of uncertainty is a major motivation for certain non-classical logics to be established. What makes NAL special on this aspect is that it attributes several types of uncertainty to a common root (AIKR), and measures them uniformly (in truth-value).

Let us see a concrete example where a medical doctor evaluates whether a patient p has a disease d.

One way to make such a judgment is to check if p has the the usual *symptoms* of d. For each symptom s_i, in NAL the inference is from $[d] \rightarrow [s_i]$ and $\{p\} \rightarrow [s_i]$ to $\{p\} \rightarrow [d]$ by *abduction*. If there are n equally weighted symptoms, and p has m of them but not the others, the n abductive conclusions can be merged by the revision rule to get a summarized conclusion $\{p\} \rightarrow [d]$ with a frequency m/n.

Here the situation is quite similar to the cases studied by fuzzy logic [Zadeh, 1965]. When the category $[d]$ (patients of d) is not defined by a sufficient and necessary condition, but specified by a set of properties (the symptoms), none of them is absolutely necessary for an instance to belong to the category. Instead, the more properties an instance has, the higher is its degree of membership, or its level of "typicalness" in the category. Using NAL terminology, we can say that here the uncertainty (i.e., m/n is usually neither 1 nor 0) comes from *the diversity of the intension* of $[d]$, since its properties s_i ($i = 1, \ldots, n$) do not describe the same set of instances.

On the other hand, the uncertainty measured by *frequency* can also come from *the diversity of the extension* of $[d]$, when some, but not all of its instances have a property. If p belongs to a *reference class* c, and among n known instances of c, m of them have d, we also have $\{p\} \rightarrow [d]$ with a frequency m/n, this time derived from $c \rightarrow [d]$ and $\{p\} \rightarrow c$ by *deduction*. This type of uncertainty is similar to the "randomness" studied in probability and statistics.

In practical situations, both the extension and intension of a concept can be diverse, so randomness and fuzziness happen together. Furthermore, each time usually only part of the extension or intension is considered, which leads to *inconsistency* in judgments, that is, $\{p\} \rightarrow [d]$ gets different truth-values when derived in different ways. In NAL, this situation is handled by the *revision* rule, and when the truth-value is supported by evidence from different sources, normally the evidence is neither purely extensional nor purely intensional, but a mixture where the two factors cannot be clearly separated.

The *confidence* value measures another type of uncertainty, *ignorance*. While *frequency* is about "positive vs. negative" (evidence), *confidence* is about "past vs. future" (evidence). Since the future is infinite in principle, only a constant amount of it (measured by k) is used to compare to the past (the known). According to the experience-grounded semantics, a higher *confidence* does not mean “closer to the truth” or “more likely to be confirmed”, but “less sensitive to new evidence”.

There are already several publications in which NAL is compared to the Bayesian approach of uncertain reasoning [Wang, 2001, Wang, 2004a, Wang, 2009]. In summary, NAL does not require the knowledge for a consistent probability distribution to be defined on all statements of interest, nor the resources for global updating when new evidence comes. Instead, it merely uses whatever knowledge and resources available to make the most reasonable judgment.

Using a numerical representation of uncertainty, NAL is more powerful than the binary logics when handling uncertainty. For instance, unlike modal logic, NAL does not divide statements into “necessarily true” and “possibly true”. Since it is open to novel experience, all empirical statements are “possibly true” to various degrees. On the other hand, analytical statements are “true within a theory”, which is not a modality. In particular, NAL rejects the notion of “possible world”, since it does not restrict its future experience to the descriptions of a given vocabulary.

Similarly, NAL is like a non-monotonic logic since it can “change its mind” in light of new evidence. However, in NAL all empirical statements are revisable, while in a non-monotonic logic the system can only change the truth-value of a *hypothesis* (e.g., “Tweety can fly”), but not that of a *default rule* (e.g., “Bird can fly”), nor a *fact* (e.g., “Tweety is a bird’). From a philosophical point of view, the separation of these three categories are not easy to establish. In NAL, their difference is a matter of degree.

3.4 Reasoning and learning

One major difference between an axiomatic system doing theorem proving and an adaptive system doing evidential reasoning is that the former is a *closed system* with fixed concepts and knowledge, while the latter is an *open system* with evolving concepts and knowledge.

NARS can start with an empty memory, and let all domain knowledge come from experience. Even though it is possible for the system to start with a preloaded memory, all judgments in it are still revisable, and there is no fundamental difference between “innate” knowledge and “acquired” knowledge. All “object-level” knowledge, e.g., knowledge expressed in Narsese, can be learned from experience, including terms with their meaning and statements with their truth-value.

In this way, “reasoning” and “learning” correspond to two different ways to describe the same process: when the running process of the system is described step by step, it is natural to be taken as reasoning, where each step follows a certain rule to derive conclusion from premises; when the focus is on the long-term effects of the process, it is natural to be taken as learning, where the system gradually acquires beliefs and concepts [Wang and Li, 2016]. On the contrary, in the mainstream AI and CogSci, reasoning and learning have been traditionally treated as two separate processes [Russell and Norvig, 2010, Wilson and Keil, 1999]; in the study of logic, learning is rarely mentioned at all.

Of course, there is still restriction on what NARS can learn. The "meta-level" knowledge of the system is innate and cannot be modified by experience. Such knowledge includes the grammar rules of Narsese, the semantic principles determining truth-value and meaning, the inference rules of NAL, etc. For the system, NAL is its "native" logic, which allows the system to learn other knowledge, including "secondary" logics, that can be applied to solve problems in a domain-specific manner. Even with the ability of self-awareness and self-control, NARS cannot fully overwrite or rewrite its own "laws of thought", though can supplement or augment them using its acquired knowledge and skills about reasoning and learning.

When designing a logic for an adaptive system, a crucial decision is on what to build into the system, and what to leave to learning. In NAL, this distinction exactly follows the "object-level vs. meta-level" line, and the meta-level is designed to be minimum while it still provides the necessary function. NAL takes the *inheritance* copula as a cornerstone, which gives the system many important features not available in other logics. On the other hand, it leaves the meaning of many other notions, such as "to believe" and "to cause", to be determined by experience, while these notions are often built-in as logical constants [Hendricks and Symons, 2015, Williamson, 2007].

One important form of learning in NAL is the learning of new concepts. In traditional and classical logics, concepts pre-exist with intrinsic meaning, though can be referred to in the logic using constant arguments or predicates. On the contrary, in NARS, a concept can be "learned" in several senses:

- When a novel term (either atomic or compound) appears in the system's experience, the system may create a concept in the memory to record the related experience if it is not already there. In this way, the system gets the concept directly from the outside.
- When a novel term is generated by a compositional or decompositional rule and there was no concept for it in the memory, it is a "creative idea" that the system learned by recognizing a novel conceptual structure. In this way, the system gets the concept from the inside.
- With the changes in the beliefs associated to a concept, the concept can gradually change its meaning within the system. When the changes are significant enough, it can be considered as a new concept, even though the term has been known to the system.

Since NAL governs all the above processes, it can also be considered as a logic of categorization [Wang and Hofstadter, 2006].

3.5 Overall evaluation of NAL

In summary, NAL is a term logic with categorical sentences and syllogistic inference rules, plus an experience-grounded semantics. NAL uniformly handles several types of uncertainty (*randomness, fuzziness, ignorance, inconsistency,* etc.) and carries out several types of inference (*deduction, induction, abduction, revision, analogy,* etc.). The reasoning process is also responsible for *learning, categorization, pattern recognition, planning,* and so on, so as to produce intelligence as a whole.

The objective of NARS, including NAL, is to provide a normative model for reasoning in *realistic* situations for all intelligent systems, including humans and computers. Accurately speaking, it is not a rival of classical logic, which regulates reasoning in highly idealized situations, such as mathematics. However, classical logic is often mistakenly applied in realistic situations, and NAL does provide an alternative for those applications.

Designed with this aim, NARS has very different properties when compared with the computational implementations of classic logic. By definition, NARS cannot have the traditionally desired properties such as *consistency* (since new knowledge may conflict with previous knowledge), *soundness* (since all predictions are fallible), *completeness* (since there are always things the system does not know), and *decidability* (since question-answering is context-sensitive, open-ended, and does not follow a fixed algorithm). Nevertheless, these properties become what the system attempts to *approach* (though never *reach*) – the system constantly resolves conflicts, corrects mistakes, absorbs new knowledge, and learns new skills. On the other hand, NARS has properties that no traditional reasoning system can possess: *adaptivity* (it revises its beliefs according to new evidence), *creativity* (it can deal with novel tasks), and *flexibility* (it manages its own resources according to the changing demands).

Of course, NAL still inherits many ideas from classical logic, as well as from Aristotelian logic, set theory, probability theory, etc., though its fundamental assumptions make it unlike an extension of any of these theories. In a broad sense, NAL is a "non-classical" logic, though its difference with classical logic is much more broader and deeper than the existing non-classical logics. NAL addresses all the issues listed previously (*uncertainty*, *relevance*, *openness*, and *amplification*) in a unified manner, by treating them as issues coming from a common root, AIKR.

NAL is not a descriptive model of human reasoning, but a normative model that is based on certain rational principles, though it is closer to a descriptive model than the other normative models. This is because its fundamental assumption is abstracted from the study of human reasoning, though when the model is designed on this assumption, the process is analytical (according to logical analysis), rather than empirical (according to psychological observations). This approach is based on the hypothesis that the principles behind human reasoning is the best solution the evolution process has found. What logic does is to express these principles in a non-biological and non-anthropocentric form, so as to become applicable into non-human systems. Such a model is not descriptive for human reasoning, because it ignores the factors that have no logical necessity in the process, though should agree with the descriptive models on the major qualitative conclusions.

NAL is not claimed to be always better than the other logical systems for all purposes. As a normative model, every logic is only applicable in situations where its fundamental assumptions are satisfied. NAL is designed for the situations where the system has insufficient knowledge and resources, with respect to the problems to be solved. For a situation where the system's knowledge and resources are sufficient, the classical logic is usually better; for a situation where the system's knowledge and resources are insufficient in certain aspect, while still sufficient in the others, some non-classical logic may be better.

NARS has the ability to learn, or even to create, a new logic for certain special situations. With respect to these "acquired logics", NAL is "built in" to the system, as a "protologic" or "logical faculty" [Hanna, 2006], and the meta-logic of the acquired logics. In the future, when a reasoning task is given, NARS can decide when to use a logic it learned, and when to use its native logic.

4 Conclusion

It is the time for logic to return to its original goal to formulate the schemes of valid inference, especially those applicable to realistic situations that intelligent systems face on a daily basis, characterized by the *insufficiency of knowledge and resources* in the systems, with respect to the tasks it must carry out.

Though classical logic has achieved great success in many domains, it was not designed for reasoning in such a situation, and often should not be applied there. Various non-classical logics have their applicable situations, but they have not moved far enough from classical logic to resolve all of its issues together, so as to provide a unified normative model of everyday reasoning.

What we need is a new logic (or new logics) that is explicitly and completely based on the assumption of insufficient knowledge and resources. Concretely, it means that the system must manage its own finite computational capacity, respect the real-time requirement associated with each problem, open to unanticipated tasks, and adapt to the changing environment. This is the normal situation in which human reasoning works, as well as the situation where AI systems are desired to work.

It is possible to redefine validity for such a situation. Though by definition in this environment it is impossible for a system to be infallible or flawless, validity can be defined here as adaptivity, which means to choose the conclusion with the strongest evidential support that the system can find using available resources. This validity can be realized by truth-preserving inference rules, when truth-value is interpreted as degree of evidential support.

To formalize such an inference process, there are reasons to believe that the proper framework is closer to Aristotle's than to Frege's, though many ideas still need to be borrowed from set theory, classical logic, non-classical logics, AI, CogSci, etc.

Implementing such a new logic in computers may eventually lead us to thinking machines. Though many criticisms to the logicist AI school are legitimate, they are usually targeting the specific type of logic used, rather than the notion of "logic" in general. The issues raised in those debates (such as those on rigidity, brittleness, and over-simplification) can be resolved by a logic like NAL.

Logic is one of the oldest branch of human wisdom, and has played crucial roles in human history. By revitalizing it, we have reason to expect it to guide the intellectual exploration again, this time both in humans and computers.

References

[Adams, 1998] Adams, E. W. (1998). *A Primer of Probability Logic*. CSLI Publications, Stanford, California.

[Anderson and Belnap, 1975] Anderson, A. R. and Belnap, N. D. (1975). *Entailment: the logic of relevance and necessity*, volume 1. Princeton University Press, Princeton, New Jersey.

[Anderson, 1990] Anderson, J. R. (1990). *The Adaptive Character of Thought*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

[Aristotle, 1882] Aristotle (1882). *The Organon, or, Logical treatises of Aristotle*. George Bell, London. Translated by O. F. Owen.

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

[Aristotle, 2005] Aristotle (2005). *Metaphysics*. NuVision Publications, LLC. Translated by W. D. Ross.

[Baader et al., 2003] Baader, F., Calvanese, D., McGuinness, D. L., Nardi, D., and Patel-Schneider, P. F., editors (2003). *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press.

[Barwise and Etchemendy, 1989] Barwise, J. and Etchemendy, J. (1989). Model-theoretic semantics. In Posner, M. I., editor, *Foundations of Cognitive Science*, pages 207–243. MIT Press, Cambridge, Massachusetts.

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

[Bocheński, 1970] Bocheński, I. M. (1970). *A History of Formal Logic*. Chelsea Publishing Company, New York. Translated and edited by I. Thomas.

[Braine and O'Brien, 1998] Braine, M. D. S. and O'Brien, D. P., editors (1998). *Mental Logic*. Lawrence Erlbaum Associates, Mahwah, New Jersey.

[Bringsjord, 2008] Bringsjord, S. (2008). The logicist manifesto: At long last let logic-based artificial intelligence become a field unto itself. *Journal of Applied Logic*, 6(4):502–525.

[Cherniak, 1986] Cherniak, C. (1986). *Minimal Rationality*. MIT Press, Cambridge, Massachusetts.

[Feigenbaum and Feldman, 1963] Feigenbaum, E. A. and Feldman, J. (1963). *Computers and Thought*. McGraw-Hill, New York.

[Flach and Kakas, 2000] Flach, P. A. and Kakas, A. C. (2000). Abductive and inductive reasoning: background and issues. In Flach, P. A. and Kakas, A. C., editors, *Abduction and Induction: Essays on their Relation and Integration*, pages 1–27. Kluwer Academic Publishers, Dordrecht.

[Frege, 1999] Frege, G. (1999). Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought. In van Heijenoort, J., editor, *Frege and Gödel: Two Fundamental Texts in Mathematical Logic*, pages 1–82. iUniverse, Lincoln, Nebraska. Originally published in 1879.

[Gabbay, 2007] Gabbay, D. M. (2007). *Logic for Artificial Intelligence and Information Technology*. College Publications, London.

[Gabbay and Woods, 2001] Gabbay, D. M. and Woods, J. (2001). The new logic. *Logic Journal of the IGPL*, 9(2):141–174.

[Gaifman, 2004] Gaifman, H. (2004). Reasoning with limited resources and assigning probabilities to arithmetical statements. *Synthese*, 140:97–119.

[Gigerenzer and Selten, 2002] Gigerenzer, G. and Selten, R., editors (2002). *Bounded Rationality: The Adaptive Toolbox*. The MIT Press.

[Goertzel, 2009] Goertzel, B. (2009). Cognitive synergy: A universal principle for feasible general intelligence? *Dynamical Psychology*.

[Goertzel et al., 2008] Goertzel, B., Iklé, M., Goertzel, I. F., and Heljakka, A. (2008). *Probabilistic Logic Networks: A Comprehensive Framework for Uncertain Inference*. Springer, New York.

[Goertzel and Pennachin, 2007] Goertzel, B. and Pennachin, C., editors (2007). *Artificial General Intelligence*. Springer, New York.

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

[Grayling, 2001] Grayling, A. C. (2001). *An Introduction to Philosophical Logic*. Wiley-Blackwell, Malden, MA, 3rd edition.

[Gust et al., 2009] Gust, H., Krumnack, U., Schwering, A., and Kühnberger, K.-U. (2009). The role of logic in AGI systems: towards a lingua franca for general intelligence. In *Proceedings of the Second Conference on Artificial General Intelligence*, pages 43–48.

[Haack, 1978] Haack, S. (1978). *Philosophy of Logics*. Cambridge University Press, Cambridge.

[Haack, 1996] Haack, S. (1996). *Deviant Logic, Fuzzy Logic: Beyond the Formalism*. University of Chicago Press, Chicago Press.

[Halpern et al., 2001] Halpern, J. Y., Harper, R., Immerman, N., Kolaitis, P. G., Vardi, M. Y., and Vianu, V. (2001). On the unusual effectiveness of logic in computer science. *The Bulletin of Symbolic Logic*, 7(2):213–236.

[Hammer et al., 2016] Hammer, P., Lofthouse, T., and Wang, P. (2016). The OpenNARS implementation of the Non-Axiomatic Reasoning System. In *Proceedings of the Ninth Conference on Artificial General Intelligence*, pages 160–170.

[Hanna, 2006] Hanna, R. (2006). *Rationality and logic*. Bradford Books. MIT Press, Cambridge, Massachusetts.

[Hayes, 1977] Hayes, P. J. (1977). In defense of logic. In *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, pages 559–565.

[Hayes, 1979] Hayes, P. J. (1979). The naïve physics manifesto. In Michie, D., editor, *Expert Systems in the Micro-Electronic Age*, pages 242–270. Edinburgh University Press, Edinburgh.

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

[Hendricks and Symons, 2015] Hendricks, V. and Symons, J. (2015). Epistemic logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Fall 2015 edition.

[Hume, 1748] Hume, D. (1748). *An Enquiry Concerning Human Understanding*. London.

[Johnson-Laird, 1983] Johnson-Laird, P. (1983). *Mental Models*. Harvard University Press, Cambridge, Massachusetts.

[Kirsh, 1991] Kirsh, D. (1991). Foundations of AI: the big issues. *Artificial Intelligence*, 47:3–30.

[Kneale and Kneale, 1962] Kneale, W. and Kneale, M. (1962). *The development of logic*. Clarendon Press, Oxford.

[Kowalski, 1979] Kowalski, R. (1979). *Logic for Problem Solving*. North Holland, New York.

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

[Lenat, 1995] Lenat, D. B. (1995). CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38.

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

[McCarthy, 1989] McCarthy, J. (1989). Artificial intelligence, logic and formalizing common sense. In Thomason, R. H., editor, *Philosophical Logic and Artificial Intelligence*, pages 161–190. Kluwer, Dordrecht.

[McDermott, 1987] McDermott, D. (1987). A critique of pure reason. *Computational Intelligence*, 3:151–160.

[Minsky, 1990] Minsky, M. (1990). Logical vs. analogical or symbolic vs. connectionist or neat vs. scruffy. In Winston, P. H. and Shellard, S. A., editors, *Artificial Intelligence at MIT, Vol. 1: Expanding Frontiers*, pages 218–243. MIT Press, Cambridge, Massachusetts.

[Montague, 1970] Montague, R. (1970). Universal grammar. *Theoria*, 36(3):373–398.

[Nilsson, 1986] Nilsson, N. J. (1986). Probabilistic logic. *Artificial Intelligence*, 28:71–87.

[Nilsson, 1991] Nilsson, N. J. (1991). Logic and artificial intelligence. *Artificial Intelligence*, 47:31–56.

[Peirce, 1931] Peirce, C. S. (1931). *Collected Papers of Charles Sanders Peirce*, volume 2. Harvard University Press, Cambridge, Massachusetts.

[Popper, 1959] Popper, K. R. (1959). *The Logic of Scientific Discovery*. Basic Books, New York.

[Priest et al., 1989] Priest, G., Routley, R., and Norman, J., editors (1989). *Paraconsistent Logic: Essays on the Inconsistent*. Philosophia Verlag, München.

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

[Russell and Norvig, 2010] Russell, S. and Norvig, P. (2010). *Artificial Intelligence: A Modern Approach*. Prentice Hall, Upper Saddle River, New Jersey, 3rd edition.

[Russell and Wefald, 1991] Russell, S. and Wefald, E. H. (1991). *Do the Right Thing: Studies in Limited Rationality*. MIT Press, Cambridge, Massachusetts.

[Schroeder-Heister, 2006] Schroeder-Heister, P. (2006). Validity concepts in proof-theoretic semantics. *Synthese*, 148:525–571.

[Simon, 1957] Simon, H. A. (1957). *Models of Man: Social and Rational*. John Wiley, New York.

[Steinkrüger, 2015] Steinkrüger, P. (2015). Aristotle's assertoric syllogistic and modern relevance logic. *Synthese*, 192:1413–1444.

[Stenning and van Lambalgen, 2008] Stenning, K. and van Lambalgen, M. (2008). *Human Reasoning and Cognitive Science*. MIT Press, Cambridge, Massachusetts.

[Turing, 1950] Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, LIX:433–460.

[Vila, 1994] Vila, L. (1994). A survey on temporal reasoning in artificial intelligence. *AI Communications*, 7(1):4–28.

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

[Wang, 2001] Wang, P. (2001). Confidence as higher-order uncertainty. In *Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications*, pages 352–361, Ithaca, New York.

[Wang, 2004a] Wang, P. (2004a). The limitation of Bayesianism. *Artificial Intelligence*, 158(1):97–106.

[Wang, 2004b] Wang, P. (2004b). Toward a unified artificial intelligence. In *Papers from the 2004 AAAI Fall Symposium on Achieving Human-Level Intelligence through Integrated Research and Systems*, pages 83–90, Washington DC.

[Wang, 2005] Wang, P. (2005). Experience-grounded semantics: a theory for intelligent systems. *Cognitive Systems Research*, 6(4):282–302.

[Wang, 2006] Wang, P. (2006). *Rigid Flexibility: The Logic of Intelligence*. Springer, Dordrecht.

[Wang, 2008] Wang, P. (2008). What do you mean by 'AI'. In *Proceedings of the First Conference on Artificial General Intelligence*, pages 362–373.

[Wang, 2009] Wang, P. (2009). Formalization of evidence: A comparative study. *Journal of Artificial General Intelligence*, 1:25–53.

[Wang, 2011] Wang, P. (2011). The assumptions on knowledge and resources in models of rationality. *International Journal of Machine Consciousness*, 3(1):193–218.

[Wang, 2013] Wang, P. (2013). *Non-Axiomatic Logic: A Model of Intelligent Reasoning*. World Scientific, Singapore.

[Wang and Goertzel, 2007] Wang, P. and Goertzel, B. (2007). Introduction: Aspects of artificial general intelligence. In Goertzel, B. and Wang, P., editors, *Advance of Artificial General Intelligence*, pages 1–16. IOS Press, Amsterdam.

[Wang and Hofstadter, 2006] Wang, P. and Hofstadter, D. (2006). A logic of categorization. *Journal of Experimental & Theoretical Artificial Intelligence*, 18(2):193–213.

[Wang and Li, 2016] Wang, P. and Li, X. (2016). Different conceptions of learning: Function approximation vs. self-organization. In *Proceedings of the Ninth Conference on Artificial General Intelligence*, pages 140–149.

[Wason and Johnson-Laird, 1972] Wason, P. C. and Johnson-Laird, P. N. (1972). *Psychology of Reasoning: Structure and Content*. Harvard University Press, Cambridge, Massachusetts.

[Whitehead and Russell, 1910] Whitehead, A. N. and Russell, B. (1910). *Principia Mathematica*. Cambridge University Press, Cambridge.

[Williamson, 2007] Williamson, J. (2007). Causality. In Gabbay, D. and Guenthner, F., editors, *Handbook of Philosophical Logic*, volume 14, pages 95–126. Springer.

[Wilson and Keil, 1999] Wilson, R. A. and Keil, F. C., editors (1999). *The MIT Encyclopedia of the Cognitive Sciences*. MIT Press, Cambridge, Massachusetts.

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

[Zadeh, 1983] Zadeh, L. A. (1983). The role of fuzzy logic in the management of uncertainty in expert systems. *Fuzzy Sets and System*, 11:199–227.