Proceedings of Machine Learning Research 192:35–47, 2022 International Workshop on Self-Supervised Learning

Intelligence: From Definition to Design

Pei Wang PEI.WANG@TEMPLE.EDU
Department of Computer and Information Sciences
Temple University, Philadelphia, PA, USA

Editor: Kristinn R. Thórisson

Abstract

This paper provides a non-technical description of the ideas behind a model of intelligence that has been formalized and computerized. These ideas are organized into a train of thought consisting of the major design decisions of the model and contrasted with the major approaches in the field of artificial intelligence. The implications of these decisions are also discussed.

Keywords: unified model of intelligence, adaptation under realistic restrictions, conceptualizing, reasoning, learning

1. Introduction

NARS (Non-Axiomatic Reasoning System) is an attempt of modeling intelligence at a level of abstraction that is suitable for both a descriptive theory of human intelligence and a normative theory of artificial intelligence. Over the decades, there have been many publications on various aspects of the model (Wang, 1986, 1995, 2006, 2013). What distinguishes this paper from the previous publications is that in the following all major ideas behind NARS are systematically summarized, without touching the formal or computational details.

The paper aims at the purposes of

- introducing NARS in an easily understandable manner,
- comparing these ideas with the common approaches,
- connecting the ideas into a coherent theory of intelligence,
- exploring the possibility of realizing these ideas in various ways.

After introducing each major idea in NARS, the related previous publications are cited where the details are described and discussed, with concrete examples and results. Given the nature of this paper, those materials are not included here, as they inevitably require more precise descriptions of NARS.

2. Objective and strategy

NARS is fully based on my working definition of intelligence as *the ability of an information-processing system to adapt to its environment while working with insufficient knowledge and resources* (Wang, 1994, 2008, 2019b).

Here the "Assumption of Insufficient Knowledge and Resources" (AIKR) is further specified as the following:

- The system must manage its *finite* processing power and storage spaces.
- The system must work in *real time*, as tasks can show up at any moment and with various time restriction (such as a deadline or decreasing utility over time).
- The system must be *open* to tasks with any content (including inconsistent information and unanticipated problems), as far as their formats are recognizable by the system.

In this definition, "to adapt" means in a life-cycle of the system, it must:

- use its past experience to guide its handling of the current situation and predicting of the future situations, even though the current and the future are usually different from the past,
- use its bounded resource supplies to meet the unbounded demands, even though the resources, especially processing time, is almost never enough.

Due to AIKR, NARS normally cannot promise absolutely correct and optimal solutions to the problems it encounters. Instead, it can be much more general, flexible, original, and adaptive than the conventional AI systems. This definition leads NARS to an objective that is different from the other more popular working definitions of intelligence (Legg and Hutter, 2007; Russell and Norvig, 2020):

- NARS attempts to abstract *intelligence* from *human intelligence* as an ability that is independent of the biological nature of the brain, as well as the brain structure formed through evolution. Even though the design of NARS is brain-inspired here or there, it does not use brain structure or neural network as an aim of approximation (Hawkins and Blakeslee, 2004).

- As an adaptive system, NARS behaves according to its own experience, which does not necessarily resemble human experience. Therefore, NARS is not aimed at becoming indistinguishable from humans in behaviors, such as passing the Turing Test (Turing, 1950). This is not because its intelligence is not comparable to that of a human, but its experience won't be sufficiently similar to a human to produce human-like behaviors.

- NARS is not designed to solve any specific application problems. It is "general purpose" in the sense that it is open to problems anticipated by neither the system itself nor its designer. It will try its best to solve such a problem, though its performance will depend on the available knowledge and resources at the moment, rather than on the system's design alone. NARS does not necessarily have human knowledge (Lenat and Feigenbaum, 1991) or skills (Nilsson, 2005), especially when implemented with a body that is very different from human.

- NARS carries out a large number of cognitive functions, such as reasoning, learning, planning, perceiving, acting, etc. However, they are basically different perspectives shown, or phenomena produced, by the same underlying process, rather than accomplished by separate modules or algorithms. Consequently, the exact form of each function is usually different from how it is specified in the current AI literature (Poole and Mackworth, 2017; Russell and Norvig, 2020).

- By my working definition, intelligence is taken to be a form of rationality, that is, it indicates what the system *should do* at various situations. However, under AIKR, what is considered to be "rational" is fundamentally different from the traditional models of rationality, such as classical logic or probability theory (Russell, 1997; Hutter, 2005). Instead, the system realizes a rationality that is *relative* to the available knowledge and resources (Wang, 2011).

Accurately speaking, NARS is proposed not as a better solution for a problem that has been studied in the AI field, but as a solution for a problem that has not been clearly recognized by the mainstream AI. This is especially the case with respect to AIKR, which, though looks natural or even trivial, has not been acknowledged by the traditional models of intelligence, nor by the techniques developed on the basis of these models. This restriction on the system's working environment shares intuition with the "bounded rationality" of Simon (1957), though is more concrete and restrictive.

Since intelligence is considered as a single principle, rather than a group of functions or capabilities that can be accomplished independently, NARS is designed to be a *unified* model, rather than a *hybrid* or *integrated* model using multiple techniques (Newell, 1990). This model starts from a minimum core that realizes the above definition of intelligence, then is gradually and incrementally extended to realize more functionalities on this foundation. Other software and hardware can be used by NARS as optional tools, rather than as necessary parts of the system (Wang, 2004c).

3. Concept-centered representation

To be adaptive under AIKR, NARS segments and abstracts its experience into recurring units, *concepts*.

In NARS, a "concept" is a data structure with a unique ID called a "term". At the interface between the system and its environment, a concept may directly correspond to a sensation obtained from a sensorimotor channel, a character received from a language channel, and so on. From them, compound terms are constructed to represent the patterns found in experience. Though it is fine to consider each term as a *symbol* representing a concept, a concept itself is not a symbol representing an external object or event. This is a fundamental difference between NARS and the traditional "symbolic AI" (Newell and Simon, 1976; McCarthy, 1989).

The experienced relations between a concept and other concepts form its *meaning* to the system. In the sensorimotor channels, the directly recognized relations are those about *time* (starting from the relative order of subsequent sensations) and *space* (starting from the relative location of concurrent sensations). Based on them, the perception-cognition process builds *implication* and *equivalence* relations among events and statements, indicating that from the occurrence of an event (or the confirmation of a statement), the same can be derived for the other. Furthermore, the same process also builds *inheritance* and *similarity* relations among concepts to organize them into a generalization hierarchy, so as to summarize experience at multiple levels of granularity and scope.

For example, given proper experience, we can expect the system to form strong and stable *inheritance* relations from "dog" to "mammal", and from "mammal" to "animal". There will also be *similarity* relations between "dog" and "cat", and between "vinegar" and "wine". A major difference between NARS and other types of ontology or knowledge graph is that all these relations have truth-values associated, so are not absolutely true. Actually, the system may also contains an *inheritance* relation from "animal" to "dog", as well as a *similarity* relation between "life" and "journey", which are not usually considered as "true" at all, though may still capture partial experience of the system in certain ways, so can be used as examples or metaphors, for instance.

The *implication* and *equivalence* relations on statements directly correspond to the inference NARS can make. From its observation, experiment, and inference, the system can summarize certain reliable succession relations among events, which can be considered as "causal relations" that allow the system to make predictions and explanations, even though the conclusions can be challenged by future experience or further consideration, under AIKR.

Beside the declarative knowledge provided by the above basic conceptual relations, NARS can also represent procedural knowledge by interpreting certain terms as executable *operators* as per logic programming (Kowalski, 1979). These operators take other terms as arguments to form *operations* that can be executed to change the (external or internal) environment in various ways. Sensation is treated as a special type of operation that acquires information from the environment according to the system's commands. Operations, including sensations, are *events* whose truth-values are evaluated in specific moments, and the basic temporal relations among events can be represented and processed, together with their implicational relations.

Under AIKR, each time a concept is used to process a task, only some of its existing relations are involved, which forms its *current* meaning. This meaning is context-sensitive, but not arbitrary or random, as it is selected from the concept's experienced relations, which forms the *general* meaning of this concept.

While the system constantly receives new terms from the environment and constructs compound terms to summarize its experience, AIKR forces it to *forget* most of them by either temporarily ignoring them or permanently removing them. The extent of remembering of a term (and the concept it identifies) is indicated by a *priority value* that summarizes the relevant factors, such as its efficiency in summarizing experience and usefulness in processing tasks. This evaluation is adjusted from time to time as the situation changes.

This model of concept (Wang and Hofstadter, 2006) is very different from the other AI systems, in which a concept either has a constant meaning (in the symbolic school) (Harnad, 1990) or is only implicitly represented by a pattern of activation or association (in the connectionist school) (Hinton et al., 1986).

Since a conceptual relation is usually an abstraction of experience, it may agree with different segments of the system's experience to different extents, that is, a statement may have both positive and negative evidence. Consequently, whether the statement is "true" becomes a matter of degree. Though multi-valued logic is not novel, what makes NARS different from various probabilistic logics (Nilsson, 1986) and fuzzy logics (Zadeh, 1983) is that its truth-value is obtained by checking the statement against the system's ever-expending experience, rather than a static description of the domain, as in model-theoretic semantics (Barwise and Etchemendy, 1989). Consequently, in NARS a second measurement is used to indicate the *confidence* of the system about current (positive vs. negative) evidential ratio, by considering the effect of future evidence (Wang, 2009b).

This numerical truth-value not only let the system indicate uncertainty in its beliefs, but also express continues measurements, such as to map a sensation of brightness into the truth-value of a statement that classifies the signal as “bright”. For more accurate representation and processing, NARS can have concepts corresponding to numbers, and process them according to acquired mathematical models.

In summary, NARS segments and abstracts its experience into *concepts* related by a few forms of substitutability. As the *meaning* of concepts and the *truth-value* of conceptual relations are defined according to the system’s own experience, NARS is fundamentally different from both the symbolic approaches and the connectionist ones in knowledge representation, and does not suffer from the “symbol grounding problem” in its original sense (Harnad, 1990), as its concepts do not become meaningful by being “grounded” in outside objects or their sensations, even though the meaning of abstract concepts can get richer with relations to sensorimotor activities that directly interacting with the environment.

4. Reasoning as concept substituting and constructing

Since adaptation under AIKR suggests a concept-centered representation of experience, while concepts and conceptual relations are symbolized by terms and statements, respectively, it is natural for NARS to be built in the framework of a reasoning system, which derives new concepts and conceptual relations from the existing ones according to a *logic*.

Logic-based AI is not a new idea at all (McCarthy, 1989; Nilsson, 1991), though it has been widely criticized as too rigid (Hofstadter, 1985; McDermott, 1987; Birnbaum, 1991). This approach has been out of favor among AI researchers, especially since the recent successes of neural networks. However, I have been arguing that the failure of the traditional “logicist AI” school is not caused by the use of *logic*, but by the *type* of logic used, which is still in the tradition of mathematical logic that was developed mainly for theorem proving in axiomatic systems (Wang, 2004a, 2019a).

This is why NARS is called “non-axiomatic”, as none of the system’s empirical knowledge has the status of an axiom, that is, with a truth-value that will not be challenged by future experience. NARS can acquire and use axiomatic subsystems (corresponding to various mathematical theories) with their own reasoning mechanisms, but they are separated from the empirical knowledge of the system (Wang, 2022a).

Under AIKR, there cannot be absolute truth, but it does not mean that every statement is equally justifiable or reasonable. According to experience-grounded semantics, in NARS the truth-value of a statement indicates how much it agrees with the available evidence, therefore, in NARS a *valid* inference rule is one that decides the truth-value of its conclusion according to the evidence provided by the premises. While the traditional notion of validity requires the inference conclusions to agree with the future experience, in an adaptive system the validity of inference can only be based on the past experience of the system, simply because that is the only thing such a system can and should depend on under AIKR.

This “validity under AIKR” provides a solution to the long-standing “Problem of Induction” that was raised by Hume (1748), who pointed out that induction (and other non-deductive inference) cannot be justified as truth-preserving because its conclusions state more than its premises. For example, even if all known ravens are black, “Ravens are black” may disagree with future observations. This is not an issue in NARS, where the truth-value of a statement only indicates its extent of agreement with the *past* experience. Consequently, induction is truth-preserving in the same sense as deduction in NARS, where all types of inference may generate unsuccessful predictions when the environment changes.

In this way, experience-grounded semantics provides a consistent justification for all the inference rules of NARS, including *deduction, induction, abduction, exemplification, revision, choice, comparison, analogy,* and more (Wang, 2013). Since the basic conceptual relations in NARS (*inheritance, similarity, implication*, and *equivalence*) indicate the substitutability of terms (in meaning) and statements (in truth-value), the inference types correspond to ways for these substitutability to transfer from the conceptual relations in the promises to that in the conclusion.

Experience-grounded semantics and the corresponding version of validity are realized more naturally in *syllogistic* rules in the tradition of *term logic* (Aristotle, 1989) than in the tradition of *predicate logic* (Frege, 1999). In a term logic, "copulas" play the same role as the basic conceptual relations in NARS, in that they are logical constants recognized in the inference rules, and their combination in the premises decides the content of the conclusion. On the contrary, though these relations can be expressed in predicate logic, they are treated there in the same way as the ordinary conceptual relations, and play no special role in the inference rules. In predicate logic and propositional logic, the inference rules only consider the truth-values of the premises and conclusion, not the conceptual relations in them.

In addition to building new relations among the existing terms (concepts), NARS also has inference rules that construct compound terms from the existing terms, using a few built-in connectors. As soon as such a compound term is constructed, a new concept identified by the term is also constructed by the system. Instead of blindly or exhaustively trying various combinations, the new compound terms are introduced to capture the perceived patterns in the system's experience, as attempts to summarize the system's experience more efficiently. These constructed terms do not need to correspond to any objectively existing objects or events, but only need to represent frequent and useful patterns in the system's experience.

5. Problem-solving by reasoning

As many other reasoning systems, NARS accepts new knowledge and answers questions according to available knowledge. Furthermore, some statements are interpreted as goals to be achieved via the execution of operations, as in logic programming (Kowalski, 1979). Overall, NARS accepts three types of tasks (or problems) from the environment:

1. new knowledge to be remembered,
2. new goals to be achieved,
3. new questions to be answered.

The first type of task is processed through *forward reasoning*, in which the new knowledge is added into the memory and interacted with the existing beliefs. The other two types of task may be directly achieved by a matching belief or operation, and are also processed through *backward reasoning* with related beliefs to produce derived tasks. In summary, a task is processed by the relevant beliefs, and this process produces derived tasks that are similarly processed recursively.

NARS needs both forward and backward reasoning. Even though in principle all of its possible conclusions can be derived using forward reasoning alone, the resource expenses are not affordable, which is why few reasoning system only depends on forward reasoning without the guidance of goals or questions. On the other hand, though some techniques (such as resolution-refutation and Prolog programs) fully depend backward reasoning, NARS needs forward inference for truth-value calculations. Therefore, a major function of backward reasoning is to activate the relevant beliefs for goals or questions.

Though the above description sounds like how ordinary reasoning systems work, AIKR changes the situation fundamentally. In ordinary systems, the processing of a task, or the solving of a problem, follows an algorithm for this type of task or problem, which is either designed by a human developer or learned by a learning algorithm from training data. In either way, the problem-specific algorithm guarantees the repeatability of the solving process and the solution, as well as the resources expense of the process (often specified as its computational complexity).

This form of "algorithmic problem-solving" is no longer feasible under AIKR. The *open* requirement means the system often needs to deal with a novel problem for which there is no known algorithm in the system; the *finite* and *real-time* requirements mean the system usually cannot exhaustively evaluate all possible solutions to find the optimal one, and even an existing algorithm may not be feasible if it takes longer time to finish than required. It is not uncommon that a problem is completely beyond the system's current ability.

Though the above situation looks harsh, it is not that different from the situations we humans have to face, and arguably where intelligence is needed. Indeed, if for a problem the system has sufficient knowledge, it means there is a known algorithm for it; if the system also has sufficient resources to meet the requirements of the algorithm, the problem is consider solved in the context of computer science, and what is left for the computer is to actually execute the solution, which hardly needs any intelligence.

To say that "intelligence is needed only when the system does not know how to solve a problem" may sound contradictory, especially when the problem is eventually solved. This is not an issue because the solution is obtained by a tentative process without the guidance of a predetermined algorithm, also the solution cannot be absolutely correct or optimal, even when it satisfies the system's need at the moment.

NARS certainly has no magic power to solve unsolvable problems. Instead, in the system the notions of "problem" and "solution" are used differently from their conventional usage in theoretical computer science. In NARS each *problem instance* is processed according to what the system knows about it, rather than by an algorithm designed for the *problem class* it belongs to. According to the currently available knowledge and resources, the system may get zero, one, or more solutions, though none of them is considered as the final solution — NARS simply reports the best it has found so far, and continue to look for better ones, like an anytime algorithm (Zilberstein, 1995). When a problem is no longer processed, it is usually not because it has been fully solved or proved unsolvable, but because it no longer gets the attention of the system. This working mode is called "case-by-case problem-solving" (Wang, 2009a).

To be adaptive under AIKR, NARS distributes its time–space resources among the existing tasks according to their relative priority values, which summarize many relevant factors and are adjusted from time to time. The objective of the allocation is not to solve any specific problem according to a fixed standard, but to achieve all existing tasks as much as possible, informally speaking. The resource allocation is strongly influenced by the system's experience, and cannot guarantee to be optimal in all future situations.

Since both the external and internal environments of the system change unpredictably, the solving process of a problem and the solution are not accurately repeatable. Consequently, notions like "algorithm", "function", and "(Turing) computation" are no longer applicable to NARS at the problem-solving (also known as task-processing) level, since there is no fixed "problem-instance to solution" mapping anymore. Similarly, at this level there is no fixed computational complexity to talk about. On the other hand, the system shows flexibility and originality, since its solution to a problem depends on the history and context of the system (Wang, 2004b).

6. Learning as self-organizing

The use of empirical concepts as the center of knowledge representation also suggests a natural and efficient memory structure for NARS, which can be approximately depicted as a network with concepts (identified by terms) as nodes and conceptual relations of predetermined types (*inheritance, similarity, implication, and equivalence*) as links. To be more specific, a conceptual relation can be a task to be processed or knowledge to be remembered, while knowledge can be further divided into beliefs and desires (Wang, 1995, 2006).

As an adaptive system, NARS spontaneously constructs and adjusts its memory to deal with the challenges from AIKR, and the process can be generally considered as "learning from its experience" and happens at different places in various forms:

Task. Every reasoning process in the system is driven by a task (new knowledge, goal, or question). All initial tasks come from the outside of the system, either as implants specified by the designer, or requests issued by the user. From the initial tasks and the available beliefs, the system recursively generates derived tasks, and these tasks work together as a "motivational complex" that decides the system's actions. This motivational complex is changed constantly be the new experience and the system's reasoning processes, though some components may gradually become stable and decide the system's long-term pursuits.

Desire. Under AIKR, the system's goals often contain conflicts in what events are desired to happen. A *desire-value* is maintained for each event to summarize positive and negative motives for its realizing, according to the system's goals. This value is gradually adjusted according to the changes outside and inside the system. Only sufficiently high desire-values trigger the execution of corresponding operations.

Belief. The system's beliefs summarize its experience by integrating information from different sources, resolving conflicting evidence, revealing hidden implications, and so on. As a result, each belief captures an aspect of the system's experience, and uses it to achieve the related tasks.

Skill. As a special case of beliefs, skills include the system's procedural knowledge, such as the preconditions and consequences of an operation. The compound operations let the system use a "program" as a single operation, even though it is formed recursively from many operations. The acquisition of skills is crucial for effective solving of complicated problems, and this also includes the skill to control certain aspects of the system's own reasoning process.

Concept. Since tasks and knowledge (beliefs and desires) all take the form of conceptual relations, they can be naturally organized around the concepts appearing in them. Since a task only directly interacts with the knowledge that shares a concept with it, all inference activities are "local" in the sense that they each happens within a concept, with the exception that at the sensorimotor front end some types of inference happen according to the temporal–spatial relations among events. This feature simplifies the implementation and enables distributed processing and hardware acceleration.

Priority. Beside adding, modifying, and removing data items (tasks, knowledge, or concepts), an important form of learning happens when the priority distribution among data items are gradually established and adjusted, also according to the system's experience. This type of "structural knowledge" is neither declarative nor procedural, but shows in the processing speed, depth, or frequency of the items.

In summary, almost all contents in the memory of NARS can be learned from the system's experience. Even though for efficiency purpose some contents can be preloaded at the beginning of a life-cycle, they are usually formed by another system (or another life-cycle of the same system) from its experience, and can still be modified when new experience comes to the current system. Consequently, whether task or knowledge is preloaded or acquired makes no fundamental difference, as the two cases can equivalently replace each other.

Though NARS can be referred to as a "learning system", it is very different from the systems developed in the current machine learning research (Flach, 2012). In NARS, "learning" is an open-ended self-organizing process that does not follow a predetermined algorithm, nor does it converge to a stable input–output mapping. Furthermore, NARS does not only learns in a bottom-up manner from concrete training samples, but can accept input at various level of abstraction, from observed sensorimotor data to communicated linguistic materials. Finally, in NARS *learning* is not a separate process integrated with reasoning and other cognitive functions, but the long-term effects of reasoning (Wang and Li, 2016; Wang, 2022b). The last conclusion applies to many other cognitive functions, too, including *planning, predicting, explaining, perceiving,* etc. — they are different aspects displayed by the same reasoning process. In this way, NARS provides a unified model of cognition (Newell, 1990; Wang, 2004c).

NARS does not learn all knowledge it has. In the system, a relatively sharp distinction is made between *object-level* and *meta-level* knowledge. The former, including the contents of memory, is learned and special-purpose (experience-bounded); the latter, including the inference rules and resource-allocation mechanisms, is built-in and general-purpose (domain-independent). Though some experience-driven parameter tuning is possible, in general the meta-level knowledge should remain constant to keep the coherence of the system within a life-cycle. It is possible to evolve the meta-level knowledge in a species of NARS, but that will be a separate process that is largely independent of the learning processes in the life-cycle of an individual system (Wang, 2007).

7. Summary

This paper provides a theory of intelligence in a condensed and informal form.

The basic ideas presented above are inspired by results of cognitive science that summarize the current knowledge about human intelligence. These results are abstracted away from their biological details and evolutionary origin, and are organized into a series of key ideas, each of which is based in the previous ones, starting from a simple working definition of intelligence.

For artificial (general) intelligence, this theory provides a unique identity that clearly distinguishes this field from other fields, such as computation theory, applied statistics, cognitive psychology, neural science, etc. This theory inherits ideas from many previous AI works, both symbolic and connectionist, but is not an extension of any of them. It provides a concrete road-map for system design that has been mostly realized in the NARS project, though many of the ideas may be implemented in other ways.

Overall, this theory aims at covering various forms of intelligence, including those of humans, animals, computers, and so on, by focusing on the principles and functions of the systems, rather than their substance, structure, or origin. The theory is not merely descriptive, but constructive and explanatory by taking intelligence as a mechanism that achieves the objective of adaptation in realistic situations.

Acknowledgments

The author would like to thank the reviewers for their helpful comments and suggestions.

References

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

Jon. Barwise and John Etchemendy. Model-theoretic semantics. In Michael I. Posner, editor, *Foundations of Cognitive Science*, pages 207–243. MIT Press, Cambridge, Massachusetts, 1989.

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

Peter Flach. *Machine Learning: The Art and Science of Algorithms That Make Sense of Data.* Cambridge University Press, New York, NY, USA, 2012.

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

Stevan Harnad. The symbol grounding problem. *Physica D*, 42:335–346, 1990.

Jeff Hawkins and Sandra Blakeslee. *On Intelligence.* Times Books, New York, 2004.

Geoffrey E. Hinton, James L. McClelland, and David E. Rumelhart. Distributed representation. In David E. Rumelhart and James L. McClelland, editors, *Parallel Distributed Processing: Exploration in the Microstructure of cognition, Vol. 1, Foundations*, pages 77–109. MIT Press, Cambridge, Massachusetts, 1986.

Douglas R. Hofstadter. Waking up from the Boolean dream, or, subcognition as computation. In *Metamagical Themas: Questing for the Essence of Mind and Pattern*, chapter 26. Basic Books, New York, 1985.

David Hume. *An Enquiry Concerning Human Understanding*. London, 1748.

Marcus Hutter. *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*. Springer, Berlin, 2005.

Robert Kowalski. *Logic for Problem Solving*. North Holland, New York, 1979.

Shane Legg and Marcus Hutter. A collection of definitions of intelligence. In Ben Goertzel and Pei Wang, editors, *Advance of Artificial General Intelligence*, pages 17–24. IOS Press, Amsterdam, 2007.

Douglas B. Lenat and Edward A. Feigenbaum. On the thresholds of knowledge. *Artificial Intelligence*, 47:185–250, 1991.

John McCarthy. Artificial intelligence, logic and formalizing common sense. In Richmond H. Thomason, editor, *Philosophical Logic and Artificial Intelligence*, pages 161–190. Kluwer, Dordrecht, 1989.

Drew McDermott. A critique of pure reason. *Computational Intelligence*, 3:151–160, 1987.

Allen Newell. *Unified Theories of Cognition*. Harvard University Press, Cambridge, Massachusetts, 1990.

Allen Newell and Herbert A. Simon. Computer science as empirical inquiry: symbols and search. *Communications of the ACM*, 19(3):113–126, 1976.

Nils J. Nilsson. Probabilistic logic. *Artificial Intelligence*, 28:71–87, 1986.

Nils J. Nilsson. Logic and artificial intelligence. *Artificial Intelligence*, 47:31–56, 1991.

Nils J. Nilsson. Human-level artificial intelligence? Be serious! *AI Magazine*, 26(4):68–75, 2005.

David L. Poole and Alan K. Mackworth. *Artificial Intelligence: Foundations of Computational Agents*. Cambridge University Press, Cambridge, 2 edition, 2017.

Stuart Russell. Rationality and intelligence. *Artificial Intelligence*, 94:57–77, 1997.

Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson, Upper Saddle River, New Jersey, 4th edition, 2020.

Herbert A. Simon. *Models of Man: Social and Rational*. John Wiley, New York, 1957.

Alan M. Turing. Computing machinery and intelligence. *Mind*, LIX:433–460, 1950.

Pei Wang. A reasoning system that can deal with uncertainty. Master's Thesis, Peking University, Beijing, 1986. [*In Chinese*].

Pei Wang. On the working definition of intelligence. Technical Report 94, Center for Research on Concepts and Cognition, Indiana University, Bloomington, Indiana, 1994.

Pei Wang. *Non-Axiomatic Reasoning System: Exploring the Essence of Intelligence*. PhD Thesis, Indiana University, 1995.

Pei Wang. Cognitive logic versus mathematical logic. In *Lecture notes of the Third International Seminar on Logic and Cognition*, Guangzhou, 2004a.

Pei Wang. Problem solving with insufficient resources. *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems*, 12(5):673–700, 2004b.

Pei Wang. 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, 2004c.

Pei Wang. *Rigid Flexibility: The Logic of Intelligence*. Springer, Dordrecht, 2006.

Pei Wang. From NARS to a thinking machine. In Ben Goertzel and Pei Wang, editors, *Advance of Artificial General Intelligence*, pages 75–93. IOS Press, Amsterdam, 2007.

Pei Wang. What do you mean by "AI". In Pei Wang, Ben Goertzel, and Stan Franklin, editors, *Proceedings of the First Conference on Artificial General Intelligence*, pages 362–373, 2008.

Pei Wang. Case-by-case problem solving. In Ben Goertzel, Pascal Hitzler, and Marcus Hutter, editors, *Proceedings of the Second Conference on Artificial General Intelligence*, pages 180–185, 2009a.

Pei Wang. Formalization of evidence: A comparative study. *Journal of Artificial General Intelligence*, 1:25–53, 2009b.

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

Pei Wang. *Non-Axiomatic Logic: A Model of Intelligent Reasoning*. World Scientific, Singapore, 2013.

Pei Wang. Toward a logic of everyday reasoning. In Jordi Vallverdú and Vincent C. Müller, editors, *Blended Cognition: The Robotic Challenge*, pages 275–302. Springer International Publishing, Cham, 2019a.

Pei Wang. On defining artificial intelligence. *Journal of Artificial General Intelligence*, 10 (2):1–37, 2019b.

Pei Wang. Axiomatic reasoning in NARS. Technical Report 15, AGI Team, Temple University, Philadelphia, USA, 2022a.

Pei Wang. A unified model of reasoning and learning. In Kristinn R. Thórisson and Paul Robertson, editors, *Proceedings of the Second International Workshop on Self-Supervised Learning*, volume 159 of *Proceedings of Machine Learning Research*, pages 28–48. PMLR, 13–14 Aug 2022b.

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

Pei Wang and Xiang Li. Different conceptions of learning: Function approximation vs. self-organization. In Bas Steunebrink, Pei Wang, and Ben Goertzel, editors, *Proceedings of the Ninth Conference on Artificial General Intelligence*, pages 140–149, 2016.

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

Shlomo Zilberstein. Operational rationality through compilation of anytime algorithms. *AI Magazine*, 16(2):79–80, 1995.