On Defining Artificial Intelligence

Pei Wang
Department of Computer and Information Sciences
Temple University
pei.wang@temple.edu
http://www.cis.temple.edu/∼pwang/

Abstract

A definitions of Artificial Intelligence (AI) associates an objective to the study, and therefore provides the cornerstone of a research paradigm. The existing working definitions are not equivalent and even incompatible, though each has its own theoretical and practical values. There are reasons to define intelligence as "adaptation with insufficient knowledge and resources", as it explains a lot of cognitive phenomena, has the potential to solve many problems, and sets a sound foundation for the field of AI.

1 The Problem

1.1 Why to define AI

It is well known that there is no widely accepted definition of Artificial Intelligence (AI) [Kirsh, 1991, Allen, 1998, Hearst and Hirsh, 2000, Brachman, 2006, Nilsson, 2009, Bhatnagar et al., 2018, Monett and Lewis, 2018]. Consequently, the label "AI" has been used with many difference senses, both within the field and outside it.

Many people do not consider it a big deal. After all, many scientific concepts get good definitions only after the research become mature, rather than at the beginning of the study. Given the complexity of intelligence, it is unrealistic to expect a commonly accepted definition of AI at the current stage. Instead of spending time in a debate on definitions, they would rather pursue whatever objective that is fruitful either in theory or in practice, no matter whether it is labeled as "AI" or not.

The above opinion is agreeable to an extent. We can neither suspend the research until a definition is accepted by the community, nor expect a consensus to be arrived merely by theoretical analysis. Nevertheless, there are still reasons to pay attention to this topic at the current time.

With the recent achievements of deep learning, AI once again becomes a hot topic that attracts a lot of public attention. The business world is making

strategies to deal with this opportunity and challenge, and there are even legal and political regulations and policies proposed to deal with "AI". However, without a clear definition of the notion, "it is difficult for policy makers to assess what AI systems will be able to do in the near future, and how the field may get there. There is no common framework to determine which kinds of AI systems are even desirable." [Bhatnagar et al., 2018]

The situation is no better within the AI community. "Theories of intelligence and the goal of Artificial Intelligence (A.I.) have been the source of much confusion both within the field and among the general public" [Monett and Lewis, 2018] – among people with different opinions on what "AI" means, there is little chance for them to agree on how to build one, or to agree on the evaluation criteria, benchmark tests, milestones, etc., which are crucial for the healthy growth of a research community. It also makes cooperation difficult among different groups.

Even for a single research project, it is common to meet "conflict of ideals", where some design decisions are based on one interpretation of "AI", which some others on a different interpretation. If these interpretations turn out to be incompatible, the project has fatal trouble that cannot be dealt with technical solutions.

This paper is a summary of my previous opinions and arguments on this topic [Wang, 1994, Wang, 2006a, Wang, 2008, Wang, 2012, Wang et al., 2018], with additional discussions to give it a more systematic and comprehensive treatment. In the following, I will start at the meta-level by discussing "definition" in general, then move to the specific case of defining intelligence and AI. After summarizing the proposed definitions, I will introduce my definition, and compare it with the others, so as to clarify some assumptions in this discussion that are often implicit or hidden.

1.2 What is a definition

In its common sense, a "definition" specifies the the meaning or significance of a word or phrase, as in dictionaries and glossaries. Even so, there are still subtle issues to be noticed in the current discussion.

First, a definition can be about either a word or a concept expressed by a word, and in most situations the debate on the "meaning of AI" is more about the latter than about the former, though it is the former that is directly mentioned. In the current discussion, if "artificial intelligence" was replaced by "computer intelligence" or "computational intelligence", the underlying problem would not change much. The same is true if the concept is expressed not in English, but another human language. After all, the key issue is not in the choice of the words, but in the idea expressed by it, so this discussion is largely langauge independent. When the concept to be expressed becomes relatively well-defined, which words are chosen to express it is still a non-trivial problem, but it is secondary. Therefore, issues like the word "artificial" may be associated to "faked" are not what I want to discuss there. Instead, the focus will be on the concepts involved.

By specifying its sufficient and necessary conditions, the definition of a concept draws its boundary, and therefore regulates its usages in thinking and communication. However, even with these obvious advantages, we cannot expect every concept to be well-defined from the very beginning, even for scientific concepts, because concepts are "fluid" in nature [Hofstadter and FARG, 1995]. For instance, the boundary of a field like physics, chemistry, and biology have been formed gradually in history, rather than according to a definition about what this field should be about at the beginning. In general, to have a clear definition is not a precondition for a concept to be used in scientific research and discussions, though it is indeed highly desired.

It is often neglected that in scientific discussions there are actually two types of definition with different properties: a "dictionary definition" is *descriptive* in the sense that it summarizes the current common usage of the concept, while a "working definition" is *prescriptive* in the sense that it specifies a desired usage of the concept. Both are useful, but for different purposes. The former can be obtained via statistics and surveys, and represents the "objective opinion", while the latter is initially proposed by a single researcher or research team, which may or may not gradually become the common opinion. Actually, a new theory often uses an existing concept in a novel way, which cannot be simply dismissed as "violating its definition".

With respect to the concept of AI, its dictionary definition is relatively clear – it is nothing but what the AI researchers have been doing. Such a definition is useful for certain purposes, such as for a journal or conference reviewer to decide whether a submission is within the scope of acceptance. On the other hand, a working definition of AI sets the research objective for an AI project – it is a clarification on "what I/we mean by *AI*", which may not agree with the dictionary definition. Given the diverse usages of the phrase at the current time, to take "what the AI researchers have been doing" as a working definition would lead a research project into chaos.

In the following, the discussion is focused on the working definition of "intelligence" mainly from the perspective of AI, rather than on its dictionary definition, though the latter is still relevant.

1.3 What is a good working definition

The task of choosing a proper working definition is not unique to AI, but is in all domains, though in most cases the choice is relatively obvious, so the decision is often simply declared, rather than justified with arguments.

One commendable exception is Carnap's treatment of the concept of "probability" [Carnap, 1950]. When attempting to provide a solid foundation for probability theory, Carnap needed to start with a proper definition of probability, or in his word, he wanted to provide an *explicatum* for the *explicandum* embedded in the common usage. Instead of simply throwing out a definition that looks good to him, he first set up the following four requirements:

1. Similarity to the explicandum,

2. Exactness,
3. Fruitfulness,
4. Simplicity.

I believe these requirements also apply to other concepts to be captured in scientific theories, including "intelligence". In the following I will discuss what each of the requirements means in the current discussion.

1.3.1 Similarity

In our terminology, this requirement asks a working definition to be similar to the dictionary definition of the concept.

Though "intelligence" has been used without a well-defined boundary (otherwise this discussion would be unnecessary), there are still some common usages that can be taken as basic, which indicate what the concept should include, and what it should exclude.

First, the concept began as an attribute of human beings, and is especially about the mental or intellectual capability displayed by human. Therefore it is historically *anthropocentric*, and if a working definition of "intelligence" could not even be applied to a normal (average) human being, it would not be acceptable – no matter how good such a definition is in other aspects, it is not about the "intelligence" as we intuitively understand, but about something else.

On the other hand, it is meaningful to talk about non-human intelligence. AI is certainly such a case, and there also have been studies on "animal intelligence" [Tomasello, 2000, Goldstein et al., 2015], "collective/group intelligence" [Leimeister, 2010, Hofstadter, 1979], and "alien/extraterrestrial intelligence" [Regis, 1985, Cabrol, 2016]. Though there are many controversies in each of these discussions, as far as we take such a discussion as meaningful, we have already accepted the usage of "intelligence" as a general concept with multiple special cases that can be different from each other here or there while still keep certain common nature, which is what "intelligence" is about [Bhatnagar et al., 2018].

According to this consideration, a working definition of "intelligence" cannot be too anthropocentric to the extent that non-human intelligence becomes impossible *by definition*. It follows that the definition cannot depend on human-specific properties, which can be biological, historical, social, etc. In intelligent being does not have to be human-like in all aspects, otherwise "intelligence" and "human intelligence" would be the same concept.

However, it does not means that we want a concept to be defined as broad as possible, as that will make it vacant and trivial. That would also violate the common usage of the concept, as people do not consider everything as intelligent. Most people do not consider a conventional computer program for sorting or arithmetic calculation as intelligent, though it does carry out certain "mental activities" and is useful and valuable.

Finally, a working definition does not need to cover all common usages of a concept. For example, in the commercial world the label "intelligent" is often

used to mean "more powerful" or "better", which is a usage that can be neglect for the current purpose, as it is not part of the "core meaning" of the concept.

1.3.2 Exactness

The demands for definition are raised partly to avoid the ambiguity of the ordinary concepts and their common usages. Ideally, a definition should provide a sufficient and necessary condition for deciding the applicability of the concept in all realistic situations.

For this reason, "intelligence" should not be defined in terms of other vague concepts, such as "mind", "thinking", "cognition", "wisdom", "consciousness", etc. without defining them first (which is no less complicated than defining intelligence). Such a definition is not wrong, but fails to draw a sharp line between intelligent beings and unintelligent ones, which is what a definition is supposed to do.

This requirement is still meaningful even if "intelligence" is considered as a matter of degree (as it should be). In this situation, the definition should provide guidance for this degree to be determined.

It is why formal definitions are preferred, as they are generally more accurate and less ambiguous. However, it should be keep in mind that since the concept of intelligence have empirical content, its definition cannot be completely formal. Furthermore, a formal definition needs interpretation when it is applied, and the existence of different interpretations may reduce the exactness of a formal definition. For example, though the mathematical meaning of "probability" is fully specified by the axioms of probability, its applications still have controversies [Carnap, 1950, Hájek, 2012].

1.3.3 Fruitfulness

This is the requirement that distinguished a working definition from a dictionary definition. When a researcher or a research team defines AI, normally it is not taken as something that already fully exists, but something to be built. To serve the role of being a research objective, a working definition of "intelligence", and the derived definition of AI, should set up a clear goal for the research, as well as to provide guidance for the following work.

Given the many faces of intelligence, there are many justifiable descriptions about it, but most of them cannot play the role of a working definition well, as they do not provide clear criteria for the design decisions when building an AI system. Of course, a definition by itself is not enough to solve all the problems in research, though it nevertheless provides the most fundamental postulations for the project. In particular, the definition distinguishes the features of human intelligence that need to be reproduced in an AI system from those that can be omitted as irrelevant.

Another function of the working definition is to shed light on the solving of the existing problems in AI. Contrary to a popular belief, in scientific research the introduction of a new concept is not encouraged, unless it contributes to

the research in a unique way. By specifying intelligence in a certain way, some traditional problems may become solvable.

Finally, a working definition of AI should given the field a proper identity by specifying its subject matter and scope, which will decide its relationship with the other fields, such as computer science and cognitive psychology. It should establish AI as a domain with its unique problems and solutions, rather than a novel label of an existing domain.

1.3.4 Simplicity

It is widely agree that a scientific concept should be as simple as possible. This requirement also appears in other forms, such as the preference of "elegance" and "beauty", which often can be interpreted as conceptual simplicity.

Though the favoring of simplicity is uncontroversial, it has been given different reasons and interpretations. Some people think that it is a self-evident principle that needs no justification; some other believe that a simpler concept is more likely to be true or correct; and some people (including me) take it as derived from the requirement of efficiency and economy of cognition – a simpler concept is just easier to use.

This requirement does not deny the complexity of the processes involving intelligence. Here the hope is to identify certain essential features of intelligence from which many other features can be implied.

1.3.5 Overall evaluation

For a given working definition, usually whether it satisfies each of the above requirement is not a matter of yes or no, but a matter of degree. It is hard to establish a general and practical way to measure the degrees, but it does not mean that they cannot lead to meaningful conclusions. Usually they are used relatively, that is, we can compare two working definitions with respect to a requirement to see which one is better. Actually this is exactly we can expect: what is needed is the best definition among the available candidates, no matter what "score" it gets.

What makes the comparison tricky is the conflicts among the requirements. It is often the case that one definition is better by one standard (say, simpler) but not as good by another one (say, less fruitful). Consequently, the final choice will be a compromise, or a "weighted sum" of the individual scores on each dimension, and the weights are decided subjectively, as different researchers value the requirements differently, though they usually agree on the relevance of them.

In conclusion, even though intelligence is hard to define, an AI researcher still inevitably gives it a working definition, as far as he/she claims to be "doing AI". The difference is only at whether the definition is carefully deliberated and explicitly announced, or implicitly accepted or revealed by the choice of research goal, approach, evaluation standard, etc. This is true even for the researchers who consider all discussions on "defining AI" to be a waste of time.

Though there is no absolute or objective standard, the working definitions are not equally good. There is indeed no widely accepted definition, but it is not a reason to accept an arbitrary one and to carry out research accordingly. On the contrary, a working definition usually decides the potentials and limitations of a research approach.

2 Practices in Defining AI

2.1 Historic development

The research goals in the field of AI have been changing over the years.

After the invention of computer in the 1940s, people soon realized that its capability is not limited to numerical calculation, and can be used to carry out many intellectual tasks that are usually considered as demanding human intelligence. Computers were called "giant electronic brains", and several visionary researchers proposed theories that stress the common features of the machines and the minds, including McCulloch and Pitts [McCulloch and Pitts, 1943], Wiener [Wiener, 1948], Shannon [Shannon and Weaver, 1949], Turing [Turing, 1950], and von Neumann [von Neumann, 1958].

Though the above researchers can be considered as pioneers of AI research, and their works have influenced generations of researchers, the research field known as AI today was mainly founded by McCarthy, Minsky, Newell, and Simon. This is not merely because they participated in the Dartmouth meeting [McCarthy et al., 1955] where the phrase of "Artificial Intelligence" was coined, but because they established three leading research centers, and their ideas have largely shaped the path of the mainstream AI for decades. The followings are their opinions about what AI is about:

> "By 'general intelligent action' we wish to indicate the same scope of intelligence as we see in human action: that in any real situation behavior appropriate to the ends of the system and adaptive to the demands of the environment can occur, within some limits of speed and complexity." [Newell and Simon, 1976]
>
> "AI is concerned with methods of achieving goals in situations in which the information available has a certain complex character. The methods that have to be used are related to the problem presented by the situation and are similar whether the problem solver is human, a Martian, or a computer program." [McCarthy, 1988]
>
> Intelligence usually means "the ability to solve hard problems". [Minsky, 1985a])

As the mainstream AI has been guided by these intuitive but vague conceptions of intelligence, it has grown into a field that lacks not only a common theoretical foundation, but also a consensus on the overall research objective. Consequently, there are many disagreements on evaluation criteria, progress milestones, benchmark problems, etc. It is normal for a scientific domain to

have competing paradigms [Kuhn, 1970], but researchers in other domains at least agree on the problems to be studied.

To the larger community of computer science and information technology, AI is usually identified by the techniques grown from it, which at different periods may include theorem proving, heuristic search, game playing, expert systems, neural networks, Bayesian networks, data mining, agents, and recently, deep learning. Since these techniques are based on very different theoretical foundations and are applicable to different problems, various subdomains have been formed within AI, such as knowledge representation, reasoning, planning, machine learning, vision, natural language processing, robotics, etc. Many researchers identify themselves much closer with these subdomains than with AI, and treat the latter like an optional label that can be added or dropped depending on the current public image of AI, which has been roller-coasting.

Different attitudes towards this diversity can be perceived from two AAAI Presidential Addresses:

> "I want to consider intelligence as a collective noun. I want to see what we in AI have thought of it and review the multiple ways in which we've conceived of it. ... to conceive of AI as the study of the design space of intelligences." [Davis, 1998]
>
> "It has been hypothesized that whatever intelligence is (and we obviously have not been able to fully define it so far), it is a multidimensional thing. ... We must consider the integration and synergies of components in an overall system to really approach some form of artificial intelligence." [Brachman, 2006]

While Davis took the diversity within the field to be an admirable feature, Brachman was concerned about the fragmentation of the research community. Though many attempts have been made in the recent years to integrate the subdomains, there is still no consensus on many major issues, including what AI or intelligence means.

2.2 Major perspectives

A recent survey [Monett and Lewis, 2018] identified hundreds of definitions of intelligence. In this section I will not analyze individual definitions, but the major perspectives from which the definitions are proposed.

As explained previously, every working definition of AI corresponds to an abstraction of the human mind. It describes the mind from a certain point of view, under the belief that it is what "intelligence" is about, and guides the construction of a computer system to "do the same".

To clarify the difference among the abstractions, in the following both humans and computers are described in a very simple formal framework, where an agent or system is specified as a tuple $\langle P, S, A\rangle$, where $P = \langle p_0, \ldots, p_t\rangle$ is the sequence of percepts acquired in a period of time (as input), $A = \langle a_0, \ldots, a_t\rangle$ is the sequence of actions executed in the period (as output), and $S = \langle s_0, \ldots, s_t\rangle$ is the

sequence of internal states the agent has gone through. When a human is written as $H = \langle P^H, S^H, A^H \rangle$ and an intelligent computer as $C = \langle P^C, S^C, A^C \rangle$, a working definition of intelligence corresponds to a way to define $C \approx H$ in terms of their components, that is, in what sense C and H are similar or even identical [Wang, 2008].

2.2.1 Structure-AI

The rationale of this perspective seems self-evident. After all, intelligence starts as notion about the mental capability produced by the human brain, so the most reliable way to reproduce it is to faithfully simulate the human brain, which is a huge neural network. Such an opinion is put in its extreme form by neuroscientists Reeke and Edelman, who argue that "the ultimate goals of AI and neuroscience are quite similar" [Reeke and Edelman, 1988].

I call this type of definition "Structure-AI", since it requires an AI system to go through isomorphic state or structure changes as the brain when they are given similar input, which will produce similar output, so the three components of the two are pairwise similar to each other:

$$P^C \approx P^H,\ S^C \approx S^H,\ A^C \approx A^H$$

From this perspective, "similar to the brain" is the main standard and justification of the design, rather than merely as a source of inspiration. For this reason, it includes the brain modeling projects [Hawkins and Blakeslee, 2004, Markram, 2006, Koene and Deca, 2013], but not the artificial neural networks [Rumelhart and McClelland, 1986].

Given the complexity of the human brain, such a project must be very difficult, and it will heavily depend on the progress of neural science, but that is not the concern of this article, as we focus on the identity it gives to AI. In that aspect, the major criticism is that this definition is too anthropocentric. As explained above, a fundamental intuition behind AI is that human intelligence is a special form of a more general "intelligence", which have other forms. Using a well-known metaphor, "to fly" and "to fly as a bird" both can be taken as engineering goals, but they are different goals. The latter is possible (though difficult), but if the former is understood as identical to the latter, many valuable designs will be omitted or even disqualified, simply because they are not similar to the original.

If it turns out to be the case that the only way to get intelligence is doing exactly what the human brain does, then AI should be considered as an ill-conceived concept. Instead, we would better talk about brain modeling or emulation. A related issue is whether "mind" can be completely reduced into "brain". If it is not the case, than a good model of the brain and a good model of the mind are not the same, and the intuitive meaning of intelligence is closer to the latter than to the former.

2.2.2 Behavior-AI

One way to acknowledge a human-like mind without demanding a human-like brain is to associate intelligence to the external behaviors of the agent. After all, if an agent behaves likes a human, it should be considered as intelligent, no matter whether it looks like a human, either inside or outside.

In the agent framework, it means that C is similar to H in the sense that

$$P^C \approx P^H,\ A^C \approx A^H$$

that is, the two should have similar input–output streams, without requiring any corresponding internal structures and states.

The best known example of this perspective is Turing Test [Turing, 1950], which states that if an computer system's verbal behaviors are indistinguishable from that of a human being, it should be considered as intelligent, or a "thinking machine".

Turing Test is intuitively appealing, and has been widely taken as the definition of AI by the public. However, within the field few project aims at pretending to be human beings. Actually, the mostly relevant works on chatbots had not be taken seriously by the mainstream until the recent years, and Turing Test has be criticized by some researchers as a distraction or even harmful [Hayes and Ford, 1995, Laird et al., 2009, Marcus et al., 2016].

The most ironic point on this matter is that Turing himself did not propose his test (he called it the "imitation game") to be the definition of thinking machines, but a *sufficient condition* of it. He explicitly acknowledged that it is not a *necessary condition* by writing "May not machines carry out something which ought to be described as thinking but which is very different from what a man does? This objection is a very strong one, but at least we can say that if, nevertheless, a machine can be constructed to play the imitation game satisfactorily, we need not be troubled by this objection." [Turing, 1950] His intention was merely to get a behavior-based standard for "thinking", though it does not have to be the only standard.

To expect an AI to behave exactly like a human is too anthropocentric to allow non-human intelligence, since human behaviors not only depends on our intellectual mechanisms, but on biological, evolutionary, and cultural factors that are unique to humans. For example, we cannot expect an extraterrestrial being to pass Turing Test, though we can expect them to be similar to us certain aspects [Minsky, 1985b].

A milder version of this perspective aims at computer models whose behaviors are *similar to* that of the human beings, though do not have to be *indistinguishable from* them. Such a project may take inspirations from psychology in its architecture or mechanisms [Newell, 1990, Franklin, 2007, Bach, 2009], or use psychological data to train a machine learning model to replicate the behavior [Flach, 2012]. Though they usually do not aim at passing Turing Test, these projects still use psychological data for evaluation and justification.

2.2.3 Capability-AI

For people whose interest in AI mainly comes from its potential practical applications, the "intelligence" of a system should be indicated by its problem-solving capability. For instance, Minsky uses the word "merely means what people usually mean—the ability to solve hard problems." [Minsky, 1985a]. It certainly makes sense, as people do judge the intelligence of each other by their problem-solving capability.

In the agent framework, it means that C is similar to H in the sense that there are moments i and j that

$$p_i^C \approx p_j^H,\ a_i^C \approx a_j^H$$

that is, the action (solution) the computer produces for a percept (problem) is similar to the action produced by a human to a similar percept. To make discussion simple, here we assume that a single percept can represent the problem, and a single action can represent the solution. In this way, the "intelligence" of a system is identified by a set of problems it can solve.

Now the question becomes: which problems requires intelligence, and which do not?

There have been various suggestions on the problems that should be considered: "There are challenge problems in planning, e-commerce, knowledge discovery from databases, robotics, game playing, and numerous competitions in aspects of natural language." [Cohen, 2005] "I suggest we replace the Turing test by something I will call the 'employment test'. To pass the employment test, AI programs must be able to perform the jobs ordinarily performed by humans. Progress toward human-level AI could then be measured by the fraction of these jobs that can be acceptably performed by machines." [Nilsson, 2005] These problems indeed have practical value, but why they need intelligence while the others do not?

One extreme position is to consider every problem-solving process as requiring intelligence, though to different extent. However, in that case AI becomes synonyms with "computer application" and "automation". "So, which parts of computer science are part of AI? We suggest a rather radical answer to this question: all of them." [Hayes and Ford, 1995] If this is the case, why do we need a new concept?

One common practice is to define AI as using computer to solve problems that are only solvable by the human mind. This answer will indeed identify certain problems among all of them, but it leads to an iconic result: as soon as a computer system is built to solve a problem successfully, the problem is no longer "only solvable by the human mind", so does not need intelligence anymore. Consequently, "AI is whatever hasn't been done yet" [Hofstadter, 1979, Schank, 1991].

Many researchers are not bothered by this situation. To them, as far as their results are valuable, whether they are labeled as "AI" does not matter. This attitude is partially responsible for the lack of a "theory of AI"—as the solving of different problems usually require different theories and techniques, there is

little hope to provide a non-trivial foundation for all of them, and consequently, AI will not become a branch of science or engineering, but a "suitcase word" [Minsky, 2006] that have no core meaning.

Another issue of defining intelligence in this way is that it conflicts with an important aspect of the common usage of the word. Intuitively, intelligence is not associated with all types of problem-solving processes, as many people still have the feeling that though today's ordinary computer systems have be able to solve many problems, the way they do so is too rigid and inflexible to be considered as intelligent. For example, people usually do not consider solving a problem by exhaustively considering each possibility to be intelligent, even though this method solves many problems perfectly. Many people intuitively associate "intelligent" with "creative", "autonomous", "flexible", and so on. Such associations are dismissed by some researchers as unrealistic expectations, but it nevertheless reveals a mismatch between what is called "AI" inside the field and the public expectation and imagination on what the research should be about.

2.2.4 Function-AI

One common way to distinguish AI from the other branches of computer science is to associate this field with the cognitive functions identified in the human mind. Currently most textbooks of AI are organized in this way, with chapters on searching, reasoning, learning, planning, perceiving, acting, communicating, etc. [Luger, 2008, Russell and Norvig, 2010, Poole and Mackworth, 2017] For each function, the typical treatment is to follow the computational paradigm: "a result in Artificial Intelligence consists of the isolation of a particular information processing problem, the formulation of a computational theory for it, the construction of an algorithm that implements it, and a practical demonstration that the algorithm is successful." [Marr, 1977]

In the agent framework, this "Function-AI" perspective takes C to similar to H in the sense that there are moments i and j that

$$a_i^C = f^C(p_i^C)\,,\ a_j^H = f^H(p_j^H)\,,\ f^C \approx f^H$$

that is, the function that maps a percept (input problem) into an action (output solution) in the computer is similar to that of a human. Since here the focus is on the functions, the actual input and output values of the two agents do not have to be similar to each other. Naturally, a system with higher intelligence should implement more such functions efficiently.

Compared to the "Structure-AI" and "Behavior-AI" discussed previously, this perspective of intelligence is less anthropocentric (though the functions are still abstracted from the human mind), and it gives the field a better identity than "Capability-AI". Even so, it has its challenges.

One issue is the fragmentation of AI [Brachman, 2006] that has been addressed previously. Since each function can be specified in isolation, there is little motivation to take the other functions into consideration, as this will complicate

the situation, and may violate the basic assumptions shared by the researchers working on the function.

Another issues is that when a function is specified in this way, it may become very different from its "natural" form in the human mind where it is tightly coupled with the other cognitive processes. One example is that in the current machine learning studies, "learning" has been commonly specified as the process of using a meta-algorithm (learning algorithm) to produce an object-level algorithm (model for a domain problem) [Flach, 2012]. This working definition is exact and simple, as well as fruitful in many domains, though is arguably only a restricted version if compared to the learning processes in the human mind [Wang and Li, 2016], even compared to the initial diverse approaches within the field [Michalski et al., 1984].

These issues have been widely recognized within the field, as shown by the calls for integration [Brachman, 2006], the notion of "AI Complete Tasks" that stresses the dependency among the functions [Shapiro, 1992], and the attempts to organize the functions into a cognitive architecture [Newell, 1990] or agent framework [Nilsson, 1998]. Even so, the problem is still far from being solved, mainly because the functions have been specified and developed according to different, even incompatible, assumptions and considerations, and therefore cannot be easily combined. This theoretical incommensurability [Kuhn, 1970] has important practical consequences, as revealed by the attempts of building integrated AI systems, where "Component development is crucial; connecting the components is more crucial" [Roland and Shiman, 2002], since the difficulties are mainly theoretical, not technical.

2.2.5 Principle-AI

As in any field, there are researchers in AI trying to find fundamental principles that can uniformly explain the relevant phenomena. Here the idea comes from the usage of "intelligence" as a form of *rationality* [Simon, 1957, Russell, 1997, Hutter, 2005, Wang, 2011] that can make the best-possible decision in various situations, according to the experience or history of the system.

In the agent framework, it means that C is similar to H in the sense that

$$A^C = F^C(P^C)\,,\ A^H = F^H(P^H)\,,\ F^C \approx F^H$$

that is, the function that maps the whole stream of percepts (experience) into the whole stream of actions (behaviors) in the computer is similar to that of a human. Again, here the focus is on the function, not the actual percepts and actions. The function is called a "principle", to stress that it is not merely about a single problem and its solution, but about the agent's life-long history in various situations, when dealing with various types of problems.

This position is widely doubted and sometimes criticized as "physics envy", as the phenomena associated with intelligence seem too complicated and heterogeneous to get a "neat" explanation [Minsky, 1990]. Until a system built according to such a definition is widely acknowledged as intelligent, most peo-

ple will not be convinced that it is possible to establish a good definition in this way.

2.2.6 Relations among the perspectives

It is well-know that the AI researchers have been taking different approaches, though these approaches have been classified differently (for example, comparing the above 5 perspectives to the 4 types in [Russell and Norvig, 2010]). A more important issue is their relationship.

The most common opinions on this matter can be expressed by the proverb "All roads lead to Rome" and the parable of "the blind men and an elephant", respectively. According to the former, the approaches are all eventually lead to the same goal, and their differences are merely caused in the paths taken; according to the latter, each approach only addresses part of the picture, and eventually they should be combined together to get a whole solution. In either way, they should be considered as complement of each other, and probably can be organized into an "atlas of intelligence" [Bhatnagar et al., 2018].

Though these opinions are not completely groundless, I think they get the situation wrong, and a more suitable metaphor is the Mount Lu in the poem of Su Shi (also known as Su Dongpo, 1037–1101): "Viewed horizontally a range; a cliff from the side; It differs as we move high or low, or far or nearby." Here Mount Lu is not the elephant described by the blind men, as the range and the cliff are not *parts* but *views* of the mountain. It is true that the previously mentioned structure, behavior, capability, function, and principle are all features of *human intelligence*, but generalizations according to each of them lead to different notions of *intelligence*, and the *artificial intelligence* systems designed accordingly are even more different, as these concepts often (though not always) require different design decisions, so it is impossible for all of them to be satisfied to the same extent in a single computer system.

It is possible for an AI project to aim at more than one researcher objectives. For example, when working on a model of mind, it will be nice if some results can find practical applications; when the direct goal is to solve a real-life problem, it may be a good idea to study how it is handled by the human mind. However, there should be an objective to be considered as primary, otherwise the project would suffer from the conflicts among the objectives.

As each definition sets a separate objective, the research paradigms established accordingly are not compatible with each other. In particular, the achieving or progressing toward one of them does not necessarily imply the same effect for another one. For example, a common belief is that brain modeling is a more fundamental approach, because as soon as the human brain is accurately simulated, human behaviors and so on will appear as consequences. This is not necessarily true, because human behaviors are not only determined by the human brain, but also by the human body and human experience, to say the least. To simulate all of those will not only be a technical challenge, but also different from the "AI" as we know it intuitively.

Therefore, accurately speaking these perspectives should be considered as

different research fields, though with overlapping parts here or there, and all called "AI" for historical reasons.

2.3 The ultimate aim of AI

A working definition of AI also (explicitly or implicitly) sets the ultimate destination for the research.

In the early years of AI research, the works was clearly targeted at computers that are generally comparable with the human mind as it was understood [Turing, 1950, McCarthy et al., 1955, Feigenbaum and Feldman, 1963]. There were ambitious projects like General Problem Solver [Newell and Simon, 1963], the Fifth-Generation Computer Systems [Feigenbaum and McCorduck, 1983], and the Strategic Computing Program [Roland and Shiman, 2002], but none of them reached their declared goal, which led to a widespread doubt about the feasibility of the "grand dream of AI", and contributed to the following "AI Winter".

To survive and to clear its name, the mainstream AI community shifted its aim to more realistic tasks, like solving practical problems and carrying out individual cognitive functions. This shift is praised as "AI becomes a science (1987–present)" [Russell and Norvig, 2002], which was later changed to "AI adopts the scientific method (1987–present)" [Russell and Norvig, 2010], because "It is now more common to build on existing theories than to propose brand-new ones, to base claims on rigorous theorems or hard experimental evidence rather than on intuition, and to show relevance to real-world applications rather than toy examples." It sounds wonderful, but is at the price of giving up the initial dream of the field. For a long time, topics like "general-purpose intelligence" and "thinking machine" became taboos, and were judged as not serious or even pseudoscience. The aim of AI had been degraded to the building of "smart tools".

For people who still believe in the original dream, this goal shifting is disappointing. In an interview with Wired in 2008, Minsky criticized AI as "brain-dead", as "Only a small community has concentrated on general intelligence".

In recent years, a renaissance has been happening in AI, partly due to the hope raised by the success of new techniques like deep learning, and partly due to the realization that the old problems cannot be sidestepped. To distinguish this types of research from the conventional works, new names have been introduced, including "Human-level AI" [Minsky et al., 2004, Nilsson, 2005, McCarthy, 2007], "Strong AI" (though not in the original meaning of [Searle, 1980]), and "Artificial General Intelligence" [Goertzel and Pennachin, 2007, Wang and Goertzel, 2007]. They carry a common message: as "AI" no longer means what it used to mean within mainstream AI, a new brand is in need for an old dream.

Though none of the new names has a commonly accepted working definition, each choice of word adding into AI does have intuitive implications and associations. The use of "human-level" suggests that the conventional AI is below the human-level; "strong" suggests that the conventional AI is "weak", and the use of "general" suggests that the conventional AI is "special purpose". Though all

of these feelings are justifiable, they provide different reasons when departing from the mainstream AI.

Intelligence is widely taken as coming with different degrees or levels. For instance, most people could agree that though certain animals should be considered as intelligent, they are at lower levels in the "ladder of intelligence" compared to the "human-level". This is the image in which "Human-level AI" are "Strong AI" are usually understood, that is, they are more advanced than the conventional AI, though they are all moving in the same direction.

"Artificial General Intelligence" (AGI) could also be interpreted in this way if a general-purpose system is nothing but many special-purpose systems combined together. However, many AGI researchers share the belief that general-purpose systems and special-purpose systems have certain fundamental differences that are qualitative, not quantitative [Wang and Goertzel, 2007]. Therefore the approaches explored in AGI are generally more unorthodox, and conventional AI is disproved not because it has not moved far enough, but because it has been moving in the wrong direction, judged by the standard of AGI.

Not only there are attempts to restore the original aim of AI (though under new names), but also attempts to get it even higher. There have been predictions about the coming of "Singularity" [Kurzweil, 2006] or "Superintelligence" [Bostrom, 2014], that is, when AI systems overtake human in the level of intelligence, they will eventually become too intelligent for us to comprehend, not mention to control. Though these notions have provided little technical predictions on how to get there, they have triggered a lot of speculations and worries.

With respect to this discussion on working definitions, both of the above notions implicitly assume that there are still many (even infinite number of) levels above the human-level in the ladder of intelligence. This assumption is not as self-evident as it seems, even if we all agree that intelligence is a matter of degree, and that human intelligence is not perfect.

Whether notions like "superintelligence" make sense depends on how "intelligence" is defined. If it is defined by a set of problems solvable by the system (as in the previously discussed "Capability-AI"), it is indeed possible, or even inevitable, for the AI systems to become "more intelligent than human", as the technical progress will increase this set for AI systems (and we have see no limit for this progress), while the set for human roughly remains the same (under usual interpretation). However, if intelligence is defined as a set of principles or "laws of thought" (as in the previously discussed "Principle-AI"), then the achieving of this kind of AI indicates that we have understood these principles well enough to implement them in computer systems, which may be more powerful than us when solving certain problems, but remain comprehensible since all the time they are following the principles we already understand.

This Principle-AI still covers systems with intelligence lower than us (as in animals or preliminary types of AI) where the principles are only partially implemented, but leaves no room for "superintelligence" or whatever it is called, as principles that are fundamentally different from intelligence and completely beyond our comprehension. However, in that case the beings would better be

called "artificial gods", rather than "artificial intelligence". This possibility cannot be logically disproved, but since it is beyond our comprehension by definition, it makes no use for us to discuss it. This respond also applies to the speculation that "AI will become so intelligent that it can improve its own intelligence" – unless "improving intelligence" gets a relatively clear interpretation, this possibility cannot be meaningfully discussed.

3 My Definition of Intelligence

3.1 Intuitions and motivations

My own opinion about the aim of AI started from the vague feeling that the traditional computational systems have a very different design principle when compared to the human mind, and this principle can explain many other differences between the machine and the mind. To be more specific, a program is traditionally designed to do something in a predetermined "correct" way, while the mind is designed to be able to "do its best" in various situations using whatever it has. Consequently, absolute correctness or optimality of solutions should not be used as the design criteria, though it is still possible to talk about what is the right thing to do in each situation, and there are guiding principles across the situations.

This opinion is obviously not novel — the ideas from the previously mentioned "Principle-AI" perspective all come from similar intuitions. The real challenge is to turn this opinion into a good working definition to guide research. It should specify in a way that is different from the other approaches (otherwise why to introduce a new definition?), while satisfying the requirements of similarity, exactness, fruitfulness, and simplicity as much as possible.

To be similar to the common usage of the word, "intelligence" should take the human mind as the most typical example, while still leave room for various types of non-human intelligence. It means the definition should not be based on human-specific features, nor to demand them to be emulated in detail, otherwise it would be a definition of human intelligence, rather than its generalization.

On the other hand, the definition cannot be so broad that the traditional computers are already considered as intelligent (though maybe at a lower level). Beside being counter-intuitive, such a definition would be redundant, as it introduces no new insight into research. Despite their great practical values, traditional computer systems have no intelligence, as they are designed according to principles that are fundamentally different from what we call intelligence. AI should not be the same as computer science. Though AI will eventually be implemented in computer systems, AI systems should show fundamental differences when compared with the traditional systems, rather than merely being able to solve more problems. Intelligence should demand a different way to design and to use computers, compared to the traditional way, which is captured by the definition of "computation".

In computer science, "computation" does not mean "whatever a computer

does", but is accurately defined as a finite and repeatable process that carries out a predetermined algorithm to realize a function mapping input data to output data [Hopcroft et al., 2007]. Roughly speaking, to solve a problem "by computation" means

1. To define the problem as a mapping from a domain of valid input values to a range of possible output values;
2. To find an algorithm that carries out this mapping step by step, starting from the given input and ending with the corresponding output;
3. To implement the algorithm in a computer system so as to use it to solve each instance of the problem.

However, this approach cannot handle a problem if any of the following is the case:

1. The problem is not "well-defined" as a mapping or function;
2. The function is well-defined, but the system has no algorithm to solve it;
3. There are implemented algorithms for the problem, but the system cannot afford the resources (mainly computational time and space) to use any of them.

It is not hard to realize that many problems handled by the human mind have these issues. We cannot solve them perfectly, though still survived and live reasonably well. Isn't this ability what "intelligence" is about? Why cannot we make computers to do the same?

The above deliberation suggests that "intelligence" is associated to a working environment and a mechanism that are both different from those of "computation". This difference is what my working definition of "intelligence" stresses, as it has grown out mainly from AI considerations.

3.2 Specification

Here is my working definition:

> Intelligence is the capacity of an information-processing system to adapt to its environment while operating with insufficient knowledge and resources. [Wang, 1995]

As explained previously, this definition contains two major points, and they specify the system's working environment and mechanism, respectively.

Here "insufficient knowledge and resources" is with respect to the concrete problems the system must deal with. In this context, "having sufficient knowledge" means the system knows the procedure of solving the problem perfectly; "having sufficient resources" means the system can afford the resources (chiefly

time) required by this procedure when applying it on the instances of the problem. To be specific, this assumed working environment requires the system to be *finite*, *real-time*, and *open*.

Being **finite** means the system's information processing capability (such as how fast it can run and how much it can remember) is roughly a constant at any period of time. This requirement seems trivial, as any concrete system is surely finite. However, to acknowledge the finite nature means the system should manage its own resources, rather than merely spending them. This requirement is mainly for the theoretical models, since most traditional ones completely ignore resource restrictions.

For the system to live and work in **real-time** means that new task of various type may show up at any moment, rather than come only when the system is idly waiting for them. It also means that every task has a response time restriction, which may be in the form of an absolute deadline, or as a relatively expressed time pressure, such as "as soon as possible". In general, we can consider the value of a solution to be a decreasing function of time, so even a correct solution may become worthless when produced too late.

Being **open** to new tasks means to make no restriction on their *content*, as long as they are expressed in an acceptable *form*. Every system surely has limitations in the signals its sensorimotor mechanism can recognize or the languages its linguistic competence can handle, but there cannot be restriction on what it can be told or asked to do. Even if new experience conflicts with its current beliefs, or the new job is beyond its current skill set, the system still should handle them reasonably.

The above three assumptions is collectively called the *Assumption of Insufficient Knowledge and Resources* (AIKR), which identifies the normal working environment of an intelligent system. Of course, this assumption is about the overall situation, not on every task, as there are surely simple tasks for which the system's knowledge and resources are relatively sufficient, at least roughly speaking. However, such tasks are not where intelligence is really demanded.

People may argue that as far as a task is processed to the system's satisfaction, it must already have the knowledge and resources from which the solution is produced. This is not really an argument against AIKR, because the insufficiency occurs at the moment when the system starts to process a task, rather than after it has been processed; moreover, the assumption is more about the overall mechanism than about the processing of an individual task. It means that the system cannot depend on predetermined procedures to process its tasks, as there are always missing or uncertain knowledge, and it does not have the time to consider every possibility when processing a task. Furthermore, the environment changes constantly, so every belief the system has at a moment may be challenged by new information.

A direct implication of AIKR is that there cannot be absolutely correct or optimal solutions. As the system is open to a unrestricted future, no prediction can have guaranteed confirmation from future observations, no matter how well it has been supported by past experiences; as the system works under a constant time pressure, it has to omit certain possibility (even though they are known to

be relevant) when processing a task, so there is always a risk of missing a better solution when a neglected possibility is taken into account.

However, it does not follow that under AIKR all strategies are equally good (or equally bad). This is where the second point in the working definition, **adaptation**, come into play. In this context, this term refers to the mechanism for a system to use its *past* experience to predict the *future* situations, and to use its *bounded* resources to meet the *unbounded* demands. As new experience becomes available, it will be absorbed into the system's beliefs, so as to put its solutions and decisions on a more stable foundation.

Though adaptation is a well-known concept and is often associated with intelligence, its usage here still contains certain subtle points:

- I consider intelligence as an advanced form of adaptation, which happens within the lifetime of a single system, and the changes it produces depend on the system's past experience. Therefore it is different from the adaptation realized via evolution in a species, where the changes happen in an experience-independent manner, then selectively kept according to the future experience, as in evolutionary computation [Holland, 1992].

- Here adaptation refers to the attempt, not the consequence. The system adjusts its behaviors according to its past experience, but that will improve the system's performance only when the future is similar to the past in the relevant aspects. Under AIKR, such a similarity can be assumed, but cannot be guaranteed in any sense (including with a probabilistic distribution). However, the future can turn out to be very different from the past, then the system's adjustments will fail to meet its anticipation, and may even make things worse. However, even in this situation, the adjustments are still considered as adaptive. In this context, whether an adjustment is adaptive is judged according to the system's past experience, rather than its future experience.

AIKR and adaptation are closely related to each other. Only when the system has insufficiency in knowledge and resources, it has the needs to adapt; on the other hand, acknowledging the insufficiency but makes no attempt to fix it is effectively equivalent to denying the insufficiency. Together they consist of a *relative rationality* [Wang, 2011], that is, to get the best allowed by the available knowledge and resources.

According to this definition, the opposite of intelligence is not "cannot solve any problem", but "having a constant and invariant ability". It is what usually called "instinct" in animals, and "computation" in computers. An intelligent system does not always work better than an instinctive (or computational) system. In a relatively static environment, AIKR can be rejected, and an instinctive system is usually simpler and more efficient. On the other extreme, in a completely unpredictable environment, neither instinct nor intelligence works. It is only in a changing but relatively stable environment (where the changes are not too fast or radical) that AIKR becomes necessary, and intelligence works better than instinct.

In this way, intelligence is defined as a methodology of problem-solving that is fundamentally different from computation, as suggested in the previous section. Of course, this does not inhibit this working definition from being used to guide the design of an AI system.

3.3 Implications

The above working definition has many implications. Here I just briefly introduce the guidance it provides in the design of NARS (Non-Axiomatic Reasoning System) [Wang, 1995, Wang, 2006b, Wang, 2013], which aims at the original goal of AI, which has been called "AGI" in recent years.

As explained above, an intelligent system defined on this way cannot always solve problems by following problem-specific algorithms, as according to AIKR, such algorithms are not always available or affordable. On the other hand, a computer system eventually runs according to algorithms. The solution of this dilemma is to use "algorithm-specified" steps to form one-time processes for each problem-instance, so as to process them in a case-by-case manner [Wang, 2009]. As the system is adaptive, its internal state changes in an acyclic way, and so does the environment, consequently the problem-solving processes are no longer accurately repeatable, and there is no algorithm for the solving of a type of problem (which is a set of problem-instances). The actual processing of a problem-instance can still be recorded and considered as an "algorithm" afterwards, but since it may not happen again, such a conception makes no contribution to the system, nor to its designers and analyzers.

Therefore, the design of such a system cannot focus on the algorithms for specific problems anymore. Instead, it should focus on the design of the algorithmic steps as the building blocks of problem solving processes, as well as the mechanism to combine these steps at the run time for each individual problem-instance. Both tasks are independent of the application domain and the specific features of the problems in the domain.

This situation naturally suggests the system to be designed in the framework of a "reasoning system", interpreted broadly. Such a system runs by repeatedly using a set of *inference rules*, each of them is specified and justified in a domain-independent manner, and these rules can be combined into *inference processes* in a flexible manner to handle various tasks. Such a system is often considered as implementing a *logic*, which specifies a knowledge representation format (often using a formal grammar), as well as the valid operations on the representation (often using formal rules). Beside the logic part, the system also needs a control part to manage the memory and to select the rule and the premises for each inference step.

This type of "logic-based AI" has been proposed and followed for a long time and by many researchers [Hayes, 1977, McCarthy, 1988, Nilsson, 1991], though it has also been widely criticized [Hofstadter, 1985, McDermott, 1987, Birnbaum, 1991] and has become much less popular in recent years. To me, the notions of "logic" and "reasoning" are still productive in AI, though the concrete logic or reasoning system built in the past for AI are not based on the

proper assumptions. According to AIKR, an intelligent system cannot "derive new truth (theorems) from given truth (axioms)" anymore, even if "true" is relaxed into "probably true" [Nilsson, 1986, Adams, 1998]. Instead, the validity of reasoning has to be justified as a form of adaptation, which leads to defining "truth-value" as the degree of evidential support [Wang, 2005], which is based on the past, though used for the future.

Though the truth-value of NARS is intuitively similar to probability, in principle it is a different measurement, as it does not follow the axioms of probability theory, by which the probability of an event (or statement) is a single number. Since the truth-value of NARS is experience-grounded, it may change as new experience comes in a way that cannot be captured by Bayesian conditioning or other methods from probability theory. This is the case partly because under AIKR, the consistency among beliefs cannot be guaranteed, though the system makes efforts to reduce the inconsistency by revising its beliefs.

This semantics does not fit predicate logics well, so NARS uses a new logic, Non-Axiomatic Logic (NAL), that is designed as a term logic, in which multiple types of inference are unified (both in format and in semantics), including deduction, induction, abduction, revision, choice, comparison, analogy, etc., in the tradition of [Aristotle, 1882] and [Peirce, 1931], though the technical details are very different. NAL also share common features with set theory, propositional logic, predicate logic, non-monotonic logic, and fuzzy logic, but still differ from them fundamentally, as none of them is designed for adaptation under AIKR.

To cover various cognitive functions, in NARS the reasoning framework is extended to include "practical reasoning", that is, reasoning on actions and goals, in a way inspired by logic programming [Kowalski, 1979]. Consequently, various cognitive functions become different aspects of the same underlying process in NARS, including learning, planning, searching, categorizing, observing, acting, communicating, etc. [Wang, 2006b, Wang, 2013]. These processes are all formulated according to the "adaptation under AIKR" principle, and only try to produce the best solution with respect to the currently available knowledge and resources.

As NARS usually processes many tasks in paralle, and new tasks come constantly both from outside (observation and communication) and inside (reasoning), under AIKR it is impossible to process each of them to its "logical end". Instead, the system distributes its resources among the tasks, biased by their priority values evaluated according to the system's experience.

For each task, its processing path depends on the related beliefs that are selected at the moment, also according to the experience of the system. Consequently, even when a task is repeated, its processing path and results may be different, as its processing context has changed. That is why it is claimed previously that at the level of task processing (or problem solving), the system's input-output relation is not a fixed mapping (or computation), and the process does not follow an algorithm (not even a randomized algorithm). Many phenomena come from this dynamic resource allocation mechanism altogether, without being simulated one-by-one: attention, forgetting, association, activation spreading, etc.

This article is not intended to serve as an introduction to NARS, and the above descriptions are used only to show that the working definition of intelligence given previously does serve as the cornerstone for the design of an A(G)I system by supporting and restricting its major design decisions, though NARS is not necessarily the only way to implement this definition in computer systems.

4 Comparison with Other Definitions

4.1 With other rational principles

The definition proposed above and the associated *relative rationality* [Wang, 2011] are clearly influenced by the *bounded rationality* proposed by Simon [Simon, 1957]. "Within the behavioral model of bounded rationality, one doesn't have to make choices that are infinitely deep in time, that encompass the whole range of human values, and in which each problem is interconnected with all the other problems in the world." [Simon, 1983]

Beside this important similarity, there are still major differences between my position and Simon's on this matter.

First, "insufficient" is more restrictive than "bounded" or "limited". Even limited knowledge and resources may still be sufficient to solve certain problems, so a trivial strategy to work with bounded rationality is to only accept tasks that fall within the range of the system's capability. On the contrary, AIKR will not allow such a strategy. Bounded rationality basically corresponds to the *finite* requirement within AIKR, while not require the system to work in real time (though it assumes limited time) or to open to novel tasks (though it assumes incomplete knowledge).

Also, Simon did not explicitly use bounded rationality to define AI, but use it mainly in the explanation of human behaviors. In his own AI projects, like GPS [Newell and Simon, 1963] and Bacon [Simon et al., 1981], the restriction of knowledge and resources was taken into consideration in certain aspects, but not in the sense of AIKR. For instance, none of these systems works in real time, and though heuristic search is used instead of exhaustive search, learning and revising of heuristic functions were not considered.

Russell and Wefald's *limited rationality* also moved in the same direction by stating that "Intelligence was intimately linked to the ability to succeed as far as possible given one's limited computational and informational resources." [Russell and Wefald, 1991] In [Russell, 1997], several types of rationality are compared, and it is argued that the closest to the needs of AI is *Bounded Optimality*, the capacity to generate maximally successful behavior given the available information and computational resources.

Russell's work had gone beyond Simon's, as it provided formal specifications of the concepts proposed. However, as its formal specification is based on decision theory and computational complexity theory, it is still not under AIKR. For example, NARS cannot solve problems merely by selecting one program from a given set of programs, but have to create programs from existing com-

ponents under a variable and unpredictable time pressure. Furthermore, as the problem-solving processes do not accurately repeat, they cannot be analyzed using computational complexity theory.

A more recent definition [Legg and Hutter, 2007] states that "Intelligence measures an agent's ability to achieve goals in a wide range of environments." This definition is formalized in the reinforcement learning framework, where "All tasks that require intelligence to be solved can naturally be formulated as a maximization of some expected utility in the framework of agents" as shown in the AIXI model of "universal intelligence" [Hutter, 2005]. The model is based on the assumption that the "true environment" can be described by a computable probability distribution unknown to the agent, whose intelligence is indicated by its ability to maximize the expected reward according to the observation so far. Among the predictions consistent with the observation, the simplest one is favored as more likely to be true.

Though AIXI shares certain intuition with NARS, their assumptions about the environment and the agent are fundamentally different.

To take the environment as a computable probability distribution means whatever the agent does, the actions can only change the rewards it gets, but cannot change the environment. It is a strong postulation about the relation between an agent and its environment, which is only justified as "in standard physics there is no law of the universe that is not computable in the above sense" [Legg and Hutter, 2007]. The problem about this justification is that the descriptions about the world provided in classical physics should not be equalized to the world itself, and this reductionist position denies the need of generalization and abstraction, which lead to descriptions of the environment that are not exactly equivalent to each other.

Another major issue is that AIXI assumes infinite computational resource, and Legg and Hutter explicitly stated that "We consider the addition of resource limitations to the definition of intelligence to be either superfluous, or wrong. ... Normally we do not judge the intelligence of something relative to the resources it uses."[Legg and Hutter, 2007]. As far as I know, all tests of intelligence have explicit or implicit time limit, and people usually do not take "to exhaustively evaluate all possibilities and pick the best" as an intelligent way of solving a problem, but that is basically what AIXI does.

To be clear, I am not claiming the Legg-Hutter definition of intelligence to be *wrong*, but at least *different* from mine and many others. NARS and AIXI target on fundamentally different problems, though both associated with the notion of intelligence. The *exactness* and *simplicity* of their definition are admirable, but it does not mean that the issues in *similarity* and *fruitfulness* can be ignored —After all, an AI theory inevitably contains empirical contents, so cannot be evaluated as a mathematical theory.

4.2 With other perspectives of AI

4.2.1 With Structure-AI

In the design of NARS, no explicit attempt has been made to simulate the brain structure, either at the whole brain scale or as a neural network. This decision does not come from considerations on *usefulness* (It will contribute greatly to neural science), *possibility* (The model will surely be more and more accurate), and *difficulty* (A scientific exploration should not be abandoned just because it is hard!), but on *generality* and *necessity*.

As argued previously, as far as we agree that "brain" and "mind" are both meaningful concepts, and "human intelligence" is a form of "intelligence", there is no strong reason to insist that the latter have to be reduced to the former when they are described or constructed in non-human systems. This is especially the case in computer systems, where the underlying physical processes are very different from biological systems, not to mention the motivational and environmental factors.

That said, the design of NARS does get many inspirations from the human brain, and there are resemblance between NARS and brain models, both in mechanism and in behavior. For example, the induction rule and comparison rule of NARS are similar to Hebbian rule in that repeated occurrence leads to substitutability, and when temporal information is added, the reasoning process can model classical conditioning [Wang and Hammer, 2015]. NARS also utilizes a forgetting mechanism to deal with the insufficiency of time and space, which makes it like the human memory [Wang, 2004].

There are reasons to believe that all types of intelligence share certain structural features, no matter whether they are biological, electronic, or something else. Even so, "to be structured as faithful to the human brain as possible" is not the objective for a system like NARS, because this requirement contains irrelevant factors with respect to what such a system aims at.

4.2.2 With Behavior-AI

For a similar reason, NARS is not designed to replicate human behaviors to the extent that will allow it to pass Turing Test, because passing such a test is not a *necessary condition* of intelligence, though it may be a *sufficient condition*. Turing was not wrong, literally speaking, but a little misleading by stressing the behavioral indistinguishability between a thinking machine and a human being, and his proposal has been misunderstood by many people.

According to my working definition of intelligence, the indistinguishability between a human mind and a thinking machine should be in the *relationship between behaviors and experience*, rather than on specific behaviors. As an AI will not have human experience, it should not behave like a human, because its behaviors should depend on the experience if itself, not that of ours. It may be so smart to the extent as being able to successfully pretend to be a human, but that should not be the only way to show its intelligence, and definitely not the most natural way.

Once again, this position does not prevent NARS from showing human-like behaviors here or there, because it is designed according to similar restrictions as imposed on the human mind when it was evolved, and it should not be a surprise that the strategies they acquired are similar, though not identical in details. For instance, several "human bias and fallacies" are reproduced in NARS, and justified as inevitable consequences of adaptation under AIKR. These phenomena are often classified improperly in the previous literature, because the normative models against them the human behaviors are judged are classical logic and probability theory, both being too idealized as they ignore the knowledge and resources restrictions in reasoning and decision making [Wang, 1996, Wang, 2001].

4.2.3 With Capability-AI

NARS is not designed to solve any specific practical problem. Instead, it aims at a theoretical (meta-)problem: how can a system learn to solve problems beyond its current capability? My definition effectively takes "intelligence" as a meta-level solution, and accordingly, an intelligent system like NARS has little innate problem-solving capability, or skills, though is equipped with the potential to acquire such skills from its experience, as far as it is not living in an environment too chaotic or adversarial to be adapted to.

There is a relatively sharp level-separation in NARS. The meta-level knowledge (including the grammar rules, inference rules, control routines, executable operations, etc.) is mostly built-in and independent to the system's experience, though there are some adjustable parameters; the object-level knowledge (including the system's beliefs, desires, goals, and skills composed recursively from the operations, etc.) is stored in the system's memory, mostly acquired from experience, and remains revisable all the time.

Therefore what domain problems NARS can solve is mostly determined by its experience (its *nurture*), not by its design (its *nature*). For a specific application, NARS should not be the choice if the designer already has an efficient design. NARS provides a better solution only when AIKR has to be acknowledged, as the other techniques are inapplicable.

In this was, NARS does not really compete with the problem-specific techniques, since they are defined for different purposes.

4.2.4 With Function-AI

As mentioned previously, NARS uniformly realizes a large number of cognitive functions studied in AI, though not as separate computational processes, but as different aspects and facets of a single process (as the the range and cliff in Su Shi's poem). Consequently, the exact form of each function is quite different in NARS compared to its conventional definition in the current AI community.

For example, "learning" is usually specified as carried out by a learning algorithm, which takes some training data as input, and produce a model learned from the data as output. The model then is used as an algorithm to solve the domain problem [Flach, 2012]. On the contrary, in NARS learning is achieved via

self-organization, which happens in all aspects at the object-level, as mentioned previously and explained in [Wang and Li, 2016]. As a result, NARS is not designed to compete with techniques like deep learning [LeCun et al., 2015], but is more flexible for situations where AIKR has to be acknowledged, since deep learning and other learning algorithms are not easily applicable there —they have trouble to learn incrementally in real-time when data comes piece-by-piece, and the objective of learning is not a single input–output mapping.

Though it sounds natural to define intelligence as a collection of cognitive functions, such a definition encourages a divide-and-conquer methodology, which is partially responsible for the current fragmentation in AI. Though the techniques developed in this way have great theoretical and practical values, they are not easy to be combined together to form a thinking machine that is comparable to the human mind in general.

4.3 With other types of intelligence

Though the working definition of intelligence proposed in this article mainly comes from AI considerations, it nevertheless covers other types of intelligence as well.

The systematic study of intelligence started in psychology, and there has been a huge literature on this topic [Gottfredson, 1997, Goldstein et al., 2015]. In general, my definition is compatible with the psychological definitions. For example, Piaget sees intelligence as "the most highly developed form of mental adaptation" [Piaget, 1960], and further stated that "Intelligence in action is, in effect, irreducible to everything that is not itself and, moreover, it appears as a total system of which one cannot conceive one part without bringing in all of it." [Piaget, 1963] Medin and Ross even have made the statement that "Much of intelligent behavior can be understood in terms of strategies for coping with too little information and too many possibilities." [Medin and Ross, 1992]. Therefore, the two major factors in my definition, adaptation and AIKR, are considered as central to intelligence by psychologists.

Even so, on the surface my definition does look different from many definitions given by psychologists, mainly because the different usages of the definitions. Every definition is introduced to draw a line and to stress a difference, so it only mentions the most important factors in the distinction it makes. In psychology, the concept of "intelligence" was introduced and is used mainly to study the difference among the intellectual capability of human beings, so the attributes shared by human beings are taken for granted, and do not need to be mentioned. AIKR is just such an attribute, as the human mind obviously works under these restrictions. However this is not necessarily the case anymore when the concept is extended to include non-human. This is why it is not a good idea for AI to directly use a psychological definition of intelligence, because here the difference and similarity to be addressed is between the human mind and the computers, while the interpersonal differences are almost negligible.

For similar reasons, it is unjustified to use human IQ tests to evaluate the intelligence level of computer systems. According to my definition, intelligence

is still a matter of degree, and one system can be more intelligent than another by being able to acquire knowledge in more forms, to reorganize in more complicated ways, or to adapt more efficiently. However, it is not necessarily testable using a fixed set of problems. Though many IQ tests are nothing but sets of problem selected according to certain consideration, it can be argued that they not only directly test the problem solving ability of the subjects, but also indirectly test their learning ability, because as the innate problem-solving abilities of human beings are quite similar, a better problem-solver must be a better learner. However, this is no longer the case for computer systems, as a system can be especially designed or trained to do well in an IQ test, while being unable to do or learn anything else.

For this reason, "human intelligence", "artificial/computer intelligence", and "intelligence" should be taken as three different concepts, with the last one to provide a proper generalization for the first two.

My definition of intelligence also covers "animal intelligence", "collective intelligence", and "extraterrestrial intelligence" as special types. For the first two types, their similarity with human intelligence is mainly in their adaptive nature, rather than in their concrete structure, behavior, capability, or function, as they can be very different from human beings in those perspectives, or those features can be derived from adaptivity. Like the case of human, every animal or group is restricted by AIKR, so it does not need to be stressed. As for extraterrestrial intelligence, my definition suggests to recognize such an entity by checking whether it can adapt to its environment, rather than by its similarity with human on other aspects.

5 Conclusion

Though it is unrealistic and unnecessary to require people to define every word they use, "intelligence" in the AI context does demand a more careful treatment. Its working definition matters, since different choices lead the research to different directions, rather than merely use a phrase differently. The current field of AI is actually a mixture of multiple research fields, each with its own goal, methods, applicable situations, etc., and they are all called "AI" mainly for historical, rather than logical, reasons.

These fields are surely related, but currently the main danger is to overlook their fundamental differences and to undiscriminatingly refer to them as "AI". This practice not only causes a lot of confusions in theoretical discussions and design processes, but also have practical consequences even for the people who do not care about theory. This is the case because to anwser any non-trivial question about AI, such as "Is AI possible?", "How to build an AI?", and "Will AI be beneficial?", the "AI" in the question must be defined or at least specified first, as different types of AI correspond to very different answers. For example, in the discussion on the "safety of AI", at least we need to clearly separate the systems whose behaviors are completely determined in its design and development phrase (its "nature") from those whose knowledge, including

moral and ethical knowledge, mainly come from its own experience after it starts to run (its "nurture"). These two types of AI cannot and should not be regulated in the same way.

According to this analysis, there is no *correct* working definition of AI, as each of them has theoretical and practical values, so is not *wrong*. However, they are not *equally good* when judged according to the criteria introduced at the beginning of this article. Though I do not expect a consensus to form soon on which one is the best, at least the ultimate incompatibility among the perspectives should be recognized.

It is still the right for each AI researcher to choose how to use the name "AI", though it should be clarified when the result is discussed, with its implications understood well. Currently many researchers are produce valuable results, but not what they desired or claimed in advance. It is well known that many important ideas and techniques initiated in AI study, though they ended up contributing to the solution of other problems. Part of the reason for this to happen is the lack of a clear understanding of the assumptions of various "AI projects".

Maybe at a future time we can find proper names for each research fields involved, so as to resolve this confusion. That will probably happen when one of the working definition of "AI" has shown undeniable success. Before that time, at least we can be more clear about what we mean by "AI", and have a relatively accurate understanding about the potentials and limitations of the concepts involved.

References

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

[Allen, 1998] Allen, J. F. (1998). AI growing up: the changes and opportunities. *AI Magazine*, 19(4):13–23.

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

[Bach, 2009] Bach, J. (2009). *Principles of Synthetic Intelligence PSI: An Architecture of Motivated Cognition*. Oxford University Press, Oxford.

[Bhatnagar et al., 2018] Bhatnagar, S. et al. (2018). Mapping intelligence: Requirements and possibilities. In Müller, V. C., editor, *Philosophy and Theory of Artificial Intelligence 2017*, pages 117–135. Springer, Berlin.

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

[Bostrom, 2014] Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, Oxford, UK, 1st edition.

[Brachman, 2006] Brachman, R. J. (2006). (AA)AI — more than the sum of its parts, 2005 AAAI Presidential Address. *AI Magazine*, 27(4):19–34.

[Cabrol, 2016] Cabrol, N. A. (2016). Alien mindscapes – a perspective on the search for extraterrestrial intelligence. *Astrobiology*, 16:661–676.

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

[Cohen, 2005] Cohen, P. R. (2005). If not Turing's Test, then what? *AI Magazine*, 26:61–67.

[Davis, 1998] Davis, R. (1998). What are intelligence? and why? 1996 AAAI Presidential Address. *AI Magazine*, 19(1):91–111.

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

[Feigenbaum and McCorduck, 1983] Feigenbaum, E. A. and McCorduck, P. (1983). *The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the world*. Addison-Wesley Publishing Company, Reading, Massachusetts.

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

[Franklin, 2007] Franklin, S. (2007). A foundational architecture for artificial general intelligence. In Goertzel, B. and Wang, P., editors, *Advance of Artificial General Intelligence*, pages 36–54. IOS Press, Amsterdam.

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

[Goldstein et al., 2015] Goldstein, S., Princiotta, D., and Naglieri, J. (2015). *Handbook of intelligence: Evolutionary theory, historical perspective, and current concepts*. Springer, New York.

[Gottfredson, 1997] Gottfredson, L. S. (1997). Mainstream science on intelligence: an editorial with 52 signatories, history, and bibliography. *Intelligence*, 24:13–23.

[Hájek, 2012] Hájek, A. (2012). Interpretations of probability. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Winter 2012 edition. URL: `http://plato.stanford.edu/archives/win2012/entries/probability-interpret/`.

[Hawkins and Blakeslee, 2004] Hawkins, J. and Blakeslee, S. (2004). *On Intelligence*. Times Books, New York.

[Hayes and Ford, 1995] Hayes, P. and Ford, K. (1995). Turing Test considered harmful. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 972–977.

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

[Hearst and Hirsh, 2000] Hearst, M. A. and Hirsh, H. (2000). AI's greatest trends and controversies. *IEEE Intelligent Systems*, pages 8–17.

[Hofstadter, 1979] Hofstadter, D. R. (1979). *Gödel, Escher, Bach: an Eternal Golden Braid*. Basic Books, New York.

[Hofstadter, 1985] Hofstadter, D. R. (1985). 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.

[Hofstadter and FARG, 1995] Hofstadter, D. R. and FARG (1995). *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. Basic Books, New York.

[Holland, 1992] Holland, J. H. (1992). *Adaptation in Natural and Artificial Systems: An Introductory Analysis With Applications to Biology, Control, and Artificial Intelligence*. MIT Press, Cambridge, Massachusetts.

[Hopcroft et al., 2007] Hopcroft, J. E., Motwani, R., and Ullman, J. D. (2007). *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, Boston, 3rd edition.

[Hutter, 2005] Hutter, M. (2005). *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*. Springer, Berlin.

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

[Koene and Deca, 2013] Koene, R. and Deca, D. (2013). Editorial: Whole brain emulation seeks to implement a mind and its general intelligence through system identification. *Journal of Artificial General Intelligence*, 4:1–9.

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

[Kuhn, 1970] Kuhn, T. S. (1970). *The Structure of Scientific Revolutions*. Chicago University Press, 2nd edition.

[Kurzweil, 2006] Kurzweil, R. (2006). *The Singularity Is Near: When Humans Transcend Biology*. Penguin Books, New York.

[Laird et al., 2009] Laird, J. E., Wray, R. E., Marinier, R. P., and Langley, P. (2009). Claims and challenges in evaluating human-level intelligent systems. In *Proceedings of the Second Conference on Artificial General Intelligence*, pages 91–96.

[LeCun et al., 2015] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep Learning. *Nature*, 521:436–444.

[Legg and Hutter, 2007] Legg, S. and Hutter, M. (2007). Universal intelligence: a definition of machine intelligence. *Minds & Machines*, 17(4):391–444.

[Leimeister, 2010] Leimeister, J. M. (2010). Collective intelligence. *Business & Information Systems Engineering*, 2(4):245?248.

[Luger, 2008] Luger, G. F. (2008). *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*. Pearson, Boston, 6 edition.

[Marcus et al., 2016] Marcus, G., Rossi, F., and Veloso, M. M. (2016). Beyond the Turing Test. *AI Magazine*, 37(1):3–4.

[Markram, 2006] Markram, H. (2006). The Blue Brain project. *Nature Reviews Neuroscience*, 7(2):153–160.

[Marr, 1977] Marr, D. (1977). Artificial intelligence: a personal view. *Artificial Intelligence*, 9:37–48.

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

[McCarthy, 2007] McCarthy, J. (2007). From here to human-level AI. *Artificial Intelligence*, 171:1174–1182.

[McCarthy et al., 1955] McCarthy, J., Minsky, M., Rochester, N., and Shannon, C. (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. URL: *http://www-formal.stanford.edu/jmc/history/dartmouth.html*.

[McCulloch and Pitts, 1943] McCulloch, W. S. and Pitts, W. H. (1943). A logical calculus of ideas immanent in neural activity. *Bulletin of Mathematical Biophysics*, 5:115–133.

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

[Medin and Ross, 1992] Medin, D. L. and Ross, B. H. (1992). *Cognitive Psychology*. Harcourt Brace Jovanovich, Fort Worth.

[Michalski et al., 1984] Michalski, R., Carbonell, J., and Mitchell, T., editors (1984). *Machine Learning: An Artificial Intelligence Approach*. Springer-Verlag.

[Minsky, 1985a] Minsky, M. (1985a). *The Society of Mind*. Simon and Schuster, New York.

[Minsky, 1985b] Minsky, M. (1985b). Why intelligent aliens will be intelligible. In Regis, E., editor, *Extraterrestrials: Science and Alien Intelligence*, pages 117–128. Cambridge University Press, Cambridge.

[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.

[Minsky, 2006] Minsky, M. (2006). *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. Simon & Schuster.

[Minsky et al., 2004] Minsky, M., Singh, P., and Sloman, A. (2004). The St. Thomas common sense symposium: designing architectures for human-level intelligence. *AI Magazine*, 25(2):113–124.

[Monett and Lewis, 2018] Monett, D. and Lewis, C. W. P. (2018). Getting clarity by defining artificial intelligence - a survey. In Müller, V. C., editor, *Philosophy and Theory of Artificial Intelligence 2017*, pages 212–214. Springer, Berlin.

[Newell, 1990] Newell, A. (1990). *Unified Theories of Cognition*. Harvard University Press, Cambridge, Massachusetts.

[Newell and Simon, 1963] Newell, A. and Simon, H. A. (1963). GPS, a program that simulates human thought. In Feigenbaum, E. A. and Feldman, J., editors, *Computers and Thought*, pages 279–293. McGraw-Hill, New York.

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

[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.

[Nilsson, 1998] Nilsson, N. J. (1998). *Artificial Intelligence: A New Synthesis*. Morgan Kaufmann, San Francisco.

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

[Nilsson, 2009] Nilsson, N. J. (2009). *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. Cambridge University Press, Cambridge.

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

[Piaget, 1960] Piaget, J. (1960). *The Psychology of Intelligence*. Littlefield, Adams & Co., Paterson, New Jersey.

[Piaget, 1963] Piaget, J. (1963). *The Origins of Intelligence in Children*. W.W. Norton & Company, Inc., New York. Translated by M. Cook.

[Poole and Mackworth, 2017] Poole, D. L. and Mackworth, A. K. (2017). *Artificial Intelligence: Foundations of Computational Agents*. Cambridge University Press, Cambridge, 2 edition.

[Reeke and Edelman, 1988] Reeke, G. N. and Edelman, G. M. (1988). Real brains and artificial intelligence. *Dædalus*, 117(1):143–173.

[Regis, 1985] Regis, E., editor (1985). *Extraterrestrials: Science and alien intelligence*.

[Roland and Shiman, 2002] Roland, A. and Shiman, P. (2002). *Strategic computing : DARPA and the quest for machine intelligence, 1983-1993*. MIT Press, Cambridge, Massachusetts.

[Rumelhart and McClelland, 1986] Rumelhart, D. E. and McClelland, J. L. (1986). PDP models and general issues in cognitive science. In Rumelhart, D. E. and McClelland, J. L., editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1, Foundations*, pages 110–146. MIT Press, Cambridge, Massachusetts.

[Russell, 1997] Russell, S. (1997). Rationality and intelligence. *Artificial Intelligence*, 94:57–77.

[Russell and Norvig, 2002] Russell, S. and Norvig, P. (2002). *Artificial Intelligence: A Modern Approach*. Prentice Hall, Upper Saddle River, New Jersey, 2nd edition.

[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.

[Schank, 1991] Schank, R. C. (1991). Where is the AI? *AI Magazine*, 12(4):38–49.

[Searle, 1980] Searle, J. (1980). Minds, brains, and programs. *The Behavioral and Brain Sciences*, 3:417–424.

[Shannon and Weaver, 1949] Shannon, C. E. and Weaver, W. (1949). *The mathematical theory of communication*. The University of Illinois Press, Urbana, IL.

[Shapiro, 1992] Shapiro, S. C. (1992). Artificial intelligence. In Shapiro, S. C., editor, *Encyclopedia of Artificial Intelligence*, pages 54–57. John Wiley, New York, 2 edition.

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

[Simon, 1983] Simon, H. A. (1983). *Reason in Human Affairs*. Stanford University Press, Stanford, California.

[Simon et al., 1981] Simon, H. A., Langley, P. W., and Bradshaw, G. L. (1981). Scientific discovery as problem solving. *Synthese*, 47:1–27.

[Tomasello, 2000] Tomasello, M. (2000). Primate cognition: Introduction to the issue. *Cognitive Science*, 24(3):351–361.

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

[von Neumann, 1958] von Neumann, J. (1958). *The Computer and the Brain*. Yale University Press, New Haven, CT.

[Wang, 1994] Wang, P. (1994). On the working definition of intelligence. Technical Report 94, Center for Research on Concepts and Cognition, Indiana University, Bloomington, Indiana.

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

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

[Wang, 2001] Wang, P. (2001). Wason's cards: what is wrong. In *Proceedings of the Third International Conference on Cognitive Science*, pages 371–375, Beijing.

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

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

[Wang, 2006a] Wang, P. (2006a). Artificial intelligence: What it is, and what it should be. In *Papers from the AAAI Spring Symposium on Between a Rock and a Hard Place: Cognitive Science Principles Meet AI-Hard Problems*, pages 97–102, Stanford, California.

[Wang, 2006b] Wang, P. (2006b). *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). Case-by-case problem solving. In *Proceedings of the Second Conference on Artificial General Intelligence*, pages 180–185.

[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, 2012] Wang, P. (2012). Theories of artificial intelligence – Meta-theoretical considerations. In Wang, P. and Goertzel, B., editors, *Theoretical Foundations of Artificial General Intelligence*, pages 305–323. Atlantis Press, Paris.

[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 Hammer, 2015] Wang, P. and Hammer, P. (2015). Issues in temporal and causal inference. In *Proceedings of the Eighth Conference on Artificial General Intelligence*, pages 208–217.

[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.

[Wang et al., 2018] Wang, P., Liu, K., and Dougherty, Q. (2018). Conceptions of artificial intelligence and singularity. *Information*, 9(4).

[Wiener, 1948] Wiener, N. (1948). *Cybernetics, or control and communication in the animal and the machine*. John Wiley & Sons, Inc., New York.