Chapter 14

NARS in the Future

As a research project, NARS has achieved many important results, as summarized in the previous chapter. However, its story is still far from complete. In this last chapter, I will briefly discuss the future plan of the project, as well as its implications.

14.1 Next steps of the project

Once again, the topics to be addressed in the project can be divided into the three levels: theory, model, and implementation (from high to low, in terms of abstraction). Usually, results on a higher level guides the work on a lower level, and results on a lower level expose problems to be solved on a higher level. Therefore, normally works on a higher level progress farther than those on a lower level, though none of the three can be finished alone.

14.1.1 NARS, by itself

As mentioned before, an implemented NARS consists of a logic part and a control part. The former includes a language and a set of inference rules, and the latter is mostly about resources allocation.

The development of the logic part of NARS has been following an incremental approach, and in each new version, only part of the formal model is added into the system, based on the part already established in the current version. Consequently, each version is more intelligent than the previous one, according to the working definition of NARS — with a richer language and more inference rules, the system is more adaptive, and more efficient in using available knowledge and resources.

After NAL-8 is implemented, the logic part of the system, the "inference engine," will be mostly finished. After that, it will be the time for massive testing, with the following tasks:

- To check the expressive power of Narsese. After NAL-8, Narsese should roughly have the expressive power of a natural language. For testing purpose, various sample texts in different domains will be manually encoded into Narsese. This task may lead to revisions in the grammar of Narsese.

- To check the inferential power of NAL. Similar to the previous task, sample human inference cases will be analyzed, and compared to what NAL will produce for the same case. NAL does not have to accurately duplicate human reasoning behavior, but the differences should be documented and explained. This task may lead to revisions in the inference rules of NAL.

Compared to the logic part, the design of the control part of the model is less mature — in the previous chapters, I draw the big picture without the details. Overall, the design of the control part is deliberately postponed until the design of the logic part is finished. This is because there is an one-way dependency between the two parts of the model. When designing the logic part, the control part can be ignored, except the general principles (such as that no conclusion can be based on all beliefs of the system, and that the same evidence should not be repeatedly used to support a conclusion, and so on). On the contrary, each time a new set of inference rules are introduced, many details of the control part have to be changed accordingly. Therefore, it does not make much sense to fine tune the control part before the logic part becomes stable.

When the logic is largely in place, it will be the time to pay more attention to control. Unlike the logic, which is mainly designed by theoretical analysis, the control part has to be designed by both theoretical analysis and empirical experimenting.

Based on the control principles introduced in the previous chapters, I will attempt to provide a more detailed model for all the resources allocation strategies and functions. It will borrow intuitions from psychology and economy, as well as from previous research in AI and cognitive sciences (such as the work on credit assignment [Holland, 1986], rational analysis [Anderson, 1990], evolutionary economics [Baum, 1998], and so on) and computer science (such as resources allocation in operating system).

The above theoretical analysis will inevitably leave certain parameters (that depend on the hardware/software host systems) to be determined by actual experimenting in the implemented system. The plan for this part is to first select a set of benchmark testing cases, with certain evaluation criteria. Then, parameters will be manually tuned to find the settings that lead to the best performance. It is very likely that for each parameter there is no "optimum value" but a "normal range," and different values in the range give systems different "personalities," and none of them is always the best in all situations. Unlike the logic part, in the control part the design will never be "done," and there will always be space for refinements. Especially, the system's capability on procedure inference may be used within the system itself, so as to NARS can achieve self-monitoring and self-control to certain extent.

14.1.2 Additional capabilities

Beyond the design of NARS *per se*, there are the following directions that future NARS-based research may explore (though I may not pursue all of them myself):

Education theory. The design of NARS, no matter how complete, only determine the initial state of the system. Though it is possible to implant some "innate beliefs" into the system, its behaviors will still inevitably be determined by its experience. Therefore, NARS as designed is like a baby that has great potential, but little built-in skill. To really make the system to accomplish any practical task, extensive "education" (or call it "training") is needed, which is nothing but external control of the system's (initial) experience. Unlike the training of current AI systems (like most connectionist models), NARS cannot be trained to "converge" to certain determined behaviors. Instead, the situation will be more like the education of human beings — the tutors will have influence, but not complete control, of the system's behaviors. Therefore, I expect the education theory for NARS to be similar (though not identical) to that for human beings.

Sensorimotor subsystem. As mentioned before, NAL-8 provides a general interface for adding sensorimotor capability into NARS. Under procedural interpretation, certain goals in the system will invoke operations defined outside NARS. These operations can be procedure calls to other software, or commands to other hardware. As results, new input tasks are presented to the system, and there are side effects within the system or in the environment. The ability to use "tools" is not required in my working definition of intelligence, but if a system has this capability, its interface language will *ipso facto* be greatly extended, and it will therefore be more intelligent than a system that has only a language interface. In this way, NARS can be customized into either an "intelligent operating system" (which can flexibly integrate all kinds of software tools for various tasks), or a "mind" of a robot with particular sensorimotor mechanism (which is not necessarily similar to that of a human being).

Special hardware. I never believed that the past failure of AI was due to von Neumann computer, and I have been building NARS on the conventional hardware/software platform. Even so, specially designed hardware will surely improve the efficiency of the system. Given the fact that all inference activities happen within individual concepts, and each inference cycle consists of several fixed steps, it is quite possible to design a special computer with multiple processors, specialized to carry out inference in parallel. However, it is important to realize that even in such a hardware, AIKR will still be true, and the system will still needs to allocate limited resources among a larger number of items and activities — the system will never have so many processors that each task will get one. Therefore, though special hardware will improve efficiency, it will not change the principles on which the system is based.

Natural-language processing. Though to be able to use a natural language is not a function of NARS by design, it is often desired for various purposes. NARS has the potential of using its general-purpose learning mechanism to learn different languages. According to my current idea, the major difference between the language processing in NARS and the current approaches is that the boundary between syntax and semantics will be broken, and linguistic concepts will be handled just like other concepts. However, given the fundamental difference between human experience and the experience of an implemented NARS, I do not expect it to use a natural language as a native speaker of the language.

Multi-system community. It will be interesting to implement multiple copies of NARS, and let them communicate (in Narsese or other languages) with each other. Due to differences in system parameters, in hardware/software, and in experience, they will develop different beliefs. On the other hand, since communication creates shared experience, we can also expect some consensus to be developed among the systems. In such a setting, we can study topics like cooperation, negotiation, speech action, game playing, self/other distinction, and so on.

System evolution. The learning capability of NARS lets the system change its belief structure according to its experience, and the self-control capability will allow the system to change its internal processes to a certain extent. Still, there are some changes that will break a system's integrity. To achieve higher intelligence, we can let a NARS community evolve, using ideas like genetic algorithm. To me, *intelligence* and *evolution* are two forms of *adaptation*. The former is experience driven and within an individual system, and the latter is experience independent and across generations. The two can be combined to achieve more complicated adaptation.

Though the details of the above tasks remain to be worked out, I have reason to believe that it makes more sense to tackle these challenges from a non-axiomatic point of view (rather than from a pure-axiomatic or semi-axiomatic point of view), since they are all closely related to adaptation under insufficient knowledge and resources. Therefore, I feel that the current NARS model constitutes a necessary step toward these goals. On the other hand, I do not include them as parts of NARS, because of the belief that they are additional features, rather than essential natures, of intelligent systems. Unlike some other AGI projects, I have always taken a "minimalist" approach when designing NARS, by only equipping the system with what I believe to be absolutely necessary for an AI system.

14.1.3 Theoretical speculations

Though in this book I have already discussed many issues in AI and cognitive sciences, there are still some important topics that may be fully addressed by the progress of NARS in the future (after the previously mentioned extensions get implemented). Though now is too early to give any conclusion to these issues, I would like to briefly speculate on how to handle them in NARS.

After NAL is fully implemented and the control mechanism of NARS becomes more complicated, the system will display certain phenomena, which are not designed into it as a separate process or function, but are produced as emergent epiphenomena produced by the mass low-level events [Hofstadter, 1979]. *Emotion* may be such an example, which corresponds to different internal modes. A certain emotion may be produced by the evaluation of the general situation of the system according to the desire values of the related statements, and lead to adjustment in resources distribution. Of course, emotions in the system will not be based on biological mechanisms, but it will serve a similar function in the system as in the human mind. It is reasonable to assume that emotion plays an important role in intelligence, but it is wrong to conclude that it cannot appear in an AI system.

As mentioned previously, NARS runs continuously, without resetting itself after a task is finished. To prevent the system from being trapped in dead-ends or ignoring unusual possibilities, it may be a good idea to periodically put the system into "rest" or even "sleep," by blocking its input channels, and reducing the activation level of the concepts and tasks. In this way, when it is "woken up," the system may try very different approaches for the old problems. Furthermore, we may even allow the system to "dream," that is, to carry out some internal activity during sleeping. Since the resources competition is much weaker in this state, the system may follow paths that are judged as too unlikely to be explored in the "sober" state. Again, though their biological functions are gone and they often cause undesired results, notions like "sleep" and "dream" (like "forget") may be found to serve important information-processing functions.

A related topic is *imagery.* After the system gets sensorimotor capability, the internal representation of sensory patterns ("mental image") will be linked to concepts as parts of their meaning, according to EGS. These links will be used in both directions. If the system has some kind of visual device, then, on one hand, a cat in the visual field will provide the sensory material for the production of a mental image within the system, which is related to the concept "cat" as the result of perception. On the other hand, the processing of the "cat" concept will activate the mental image (as part of its meaning). In this way, "reasoning by image" may become just a special case of the general "reasoning by concept" process.

The "task alienation" phenomena in NARS may produce many interesting results. For example, we may find the system carries out certain activities that are not directly related to any input task, and seem have little practical utility. The system seems to be doing it "just for fun," like when humans play games. Similarly, the system may prefer certain sensation/perception patterns and processes, while the information perceived has little practical utility, like when human beings enjoy various types of art. If we analyze these processes step by step from the very beginning, I believe that originally they do serve other purposes. However, with insufficient knowledge and resources, derived tasks gradually become independent of the original ones. This tendency will be reinforced by the socialization process in multi-system society, where a system may learn to pursue a goal without knowing why, except that it is pursued by the other systems.

After NAL-8 is fully implemented, the system can be given certain sensorimotor capability, through it the system interacts with its environment directly (without using Narsese). Here the "environment" can be the inside of the system itself, which means the system will have certain self-monitoring and self-control capability. At this point, we may begin to touch the notion of *consciousness*. In principle, internally-oriented sensorimotor will be just like externally-oriented, except with very different "sensors" and "operators". The system will be able to answer questions like "What are you thinking?" according to what is actually happening in the system (rather than according to how people usually answer this question, like many "chatbots" do). Given limited knowledge and resources, the self perception of the system cannot be complete — there are lots of "subconscious" processes going on, beyond its vision. It will get ideas that seem pop up from nowhere ("inspirations"), and beliefs that cannot be traced back to their supporting evidence ("intuitions"). The system may also have its *mind-body problem*, because its internally-oriented and externally-oriented sensorimotor mechanisms will correspond to different "vocabulary" and therefore lead to different categorization. When the system attempts to describe the cause or effect of an internal process according to an external observation, it will need to cross the gap between the two, which is not always easy.

Finally, related to the above topics are the moral and ethical issues of AI research. Since the tasks (or call them motivations, drives) of NARS are determined both by the initial design (there may be built-in tasks) and by the system's experience, it will not necessarily generate "evil" ones, such as "to dominate the world." On the other hand, it will also not necessarily be friendly to human beings. Like other technology, AI is "morally neutral," and can either lead to good results or bad ones. In this sense, it is not more dangerous than other technologies. Given its flexibility, there is no way to put some foolproof safety device into NARS to prevent bad results. Our attitude toward AI should be the same as toward all scientific and technical research goals, that is, to explore them carefully, and to predict their practical consequences (though we can never be absolutely certain). Since I currently see much more reason to continue my research than reason to stop it, I will keep going on.

14.2 What NARS is not

Even after all the extensions and refinements mentioned above, there are still things that NARS cannot do, simply because it is not designed for them.

14.2.1 NARS and the other AI schools

Obviously, NARS will not reach the aims set up by the other AI schools.

NARS is not designed to simulate the human brain. I still believe that what we call "intelligence" (and the related notions like "thinking," "mind," "cognition," and so on) can be abstracted from their realization in the human brain. Of course, function and structure are not completely independent of each other, so if certain aspects of NARS turn out to be similar to what is observed in the human brain, it is not a coincidence. For example, it is not hard to recognize the similarity among the induction rule of NARS, Hebbian Learning, and Pavlovian Conditioning.

NARS is not designed to be an accurate model of human reasoning. The system should follow the same *principles* as does the human mind. However, it is not necessary to have the same internal *structure* and *mechanisms* as in the human mind, since computer hardware is fundamentally different from human wetware. Moreover, since NARS' experience will always be different from that of a human being, it is not necessary (though it is still possible to a certain extent) to have the same external behaviors as the human mind, such as exactly reproducing some psychological data or passing the Turing Test.

Though NARS does have great potential for various applications where no existing technique can be used (due to insufficient knowledge and resources), it is not designed to solve particular domain problems. When NARS is used to solve practical problems, it cannot guarantee that its results will be *correct* or *optimal* (in the eye of an omniscient observer); judgments in NARS are always subject to being revised by the system or refuted by future experience. For any given problem, it is always possible to design a special-purpose system that works better than NARS. It is like the relation between hands and tools: for any given job, it is always possible to design a tool that works better than our bare hands. However, I will not trade my hands for any tool, because of their generality and flexibility. Of course, a hand/tool combination is better than either of the two alone. For the same reason, for a given problem, it is better to let NARS use a special "tool" (i.e., special-purpose software/hardware) if it is available, rather than directly handle the problem by the system. However, I will not build these tools into NARS, just like I will not implant a hammer into my hands.

As discussed in the previous chapters, NARS has many cognitive functions, but they are usually specified quite differently from those in the other AI projects. We have discussed such cases as "learning," "induction," "planning," and so on. In NARS, they are interwoven processes without clear-cut ending points, rather than algorithms that generate certain output from certain input. Therefore, accurately speaking, what makes NARS different from many other AI projects is not the *solutions* it provides, but the *problems* it aims at. NARS is not aimed at any of those traditional AI problems, because I don't think those problems are actually related to the essence of intelligence. To me, in terms of the notion introduced in this book, they are "alienated subtasks" of AI that are derived from the original AI problem through biased and distorted beliefs.

Since NARS works with insufficient knowledge and resources, it is impossible for it to have properties that only a pure-axiomatic system can have, such as *consistency*, *completeness*, *decidability*, and so on (though the system does its best to move toward these aims when it is running). At the same time, the system is not bounded by the restrictions of pure-axiomatic systems, neither. For example, Gödel's Theorem is not directly applicable to NARS. Some people claim that Gödel's Theorem has set the limitation of AI [Lucas, 1961, Penrose, 1994], by implicitly assuming that an AI system has to be consistent. To me, the situation is just the opposite (that is, an intelligent system must have internal contradictions and conflicts), and therefore their conclusions are wrong.

The above "limitations" of NARS are easy to deal with — we can just ignore them. This is not to say that the attempt to overcome them is not a valuable goal for research, but simply that such a goal is fundamentally different from (though still related to) my current goal — exploring the essence of intelligence. These limitations of the NARS project mean that if someone is looking for a computer model with these properties, then NARS should not be a candidate, having been designed with other goals in mind.

14.2.2 How to criticize or reject NARS

As a scientific research project, the theory, model, and system of NARS can all be criticized or rejected, given valid evidence and argument.

Unfortunately, most of the criticisms I received so far are invalid, because they try to evaluate NARS in a school that it does not belong to. According to the previous discussion, it should be clear that NARS cannot be criticized for "being biologically unrealistic," "cannot pass the Turing Test," "hasn't solved any practical problem," "not working on the problems as formally specified by the AI community," "often making mistakes," "cannot find an optimum solution," and so on.

For such a system, what are valid ways to criticize it? The following is a list for a critical reader, that goes along the logical path of this book:

- You can challenge the four criteria (borrowed from Carnap) on good working definition.
- You can suggest a better working definition of intelligence, according to the above criteria.
- You can argue that a reasoning system is not the best way to formalize my working definition of intelligence.
- You can propose better selections for the components of the formal model (its formal language, semantics, inference rules, memory structure, control strategy, and so on) to implement my working definition of intelligence.
- You can design a computer system that implements the formal model in a better way.
- You can identify inconsistency among my descriptions and discussions about NARS.

Even if a valid criticism of NARS is accepted, it does not mean that the theory/model/system has been "falsified," [Popper, 1959] and should be completely rejected. Instead, it usually leads to a revision of NARS.

When will I give up this project, and move to a completely different approach? It will happen only when that approach works better in general in explaining human cognition and producing a thinking machine.

Until such a moment, the NARS project will be continued.

14.3 General implications

Finally, I will briefly mention some general implications of this research.

14.3.1 For AI

Wolfram said in an interview that “I’m convinced that after it’s understood, it really won’t be difficult to make artificial intelligence. It’s just that people have been studying absolutely the wrong things in trying to get it.”[Stork, 1997a] I quite agree with him on this comment (though we have very different opinions on what is the “right thing” to do).

From what was described in this book, it can be observed that there is nothing complicated or fancy in the technical details of NARS. However, the philosophy and methodology of this research is quite different from, and sometimes even exactly opposite to, what is accepted by the mainstream AI. It is mainly on “what to do,” rather than on “how to do it” that NARS differs from the other AI projects, and this partially explains why it is hard for this research to get accepted by the AI community.

Though the problem of AI is still not completely solved, I believe that I have given it a better clarification, and that my work finished so far has shown the potential of this research direction. Especially, it is shown to be possible to treat many problems according to the same theoretical foundation, and in the same formal model. In this way, we can give AI its identity as a scientific discipline.

As a science, this theory can explain many phenomena in human thinking. Especially, it will provide a unified picture about how the mind works and why it does not work in another way.

As a technology, NARS does not really compete with existing computer techniques, but is aiming at a domain where no current computer system can be used.

14.3.2 For cognitive sciences

Though NARS is mainly a research project in AI, its results nevertheless have contributions to other disciplines in cognitive sciences.

The most directly related discipline is *logic*. NAL can be seen as an attempt to move the subject matter of logic from the foundation of mathematics back to the regularity in thinking. In this sense, it is just the reverse of what Frege did when founding mathematical logic.

In this book, I have addressed several issues in *psychology*, such as in the previous study of reasoning, learning, categorization, human errors, and so on. Designed as a normative model, NARS does not compete with the descriptive models of aspects of human cognition. Instead, it reveals some misunderstanding and confusion in psychology when normative models are related to psychological observations.

At the current stage, the contribution of NARS to *linguistics* is mainly in the field of semantics. I believe that EGS can be used to explain many linguistic phenomena. In the future, the research may produce more results on syntax, pragmatics, and other fields.

In this book, I have discussed many topics in *philosophy*, especially, in philosophy of mind (how the mind works), philosophy of logic (what logic is about), and philosophy of science (what is a good scientific theory). Since every philosophical theory is based on certain opinions about intelligence/mind/thinking, we can expect a revolution in philosophy when the AI problem is finally solved. AI may have a larger impact in philosophy than the ones caused by Newton, Darwin, and Einstein.

Part of this project is the attempt to establish a scientific theory of intelligence. It is my opinion that a "scientific theory" is nothing but a system of shared beliefs and tasks in a community, and it is formed and changed according to the same principle of intelligence as those of an individual system, as discussed in this book. It follows that a good theory should have the following properties:

objective. A good theory should be consistent with the experience of the individuals in the community (so it cannot merely be idiosyncratic opinions),

structured. A good theory should summarize the beliefs into a simple knowledge structure (so it cannot just be a collection of unrelated judgments),

instructive. A good theory should provide concrete predictions for the future (so it cannot be vague, ambiguous, vacuous, or tautological).

The theory about intelligence presented in this book is developed to meet these criteria.

Eventually, I believe that this research will lead us to a general theory on intelligence, which covers the fields of human intelligence, computer intelligence, animal intelligence, group intelligence, extraterrestrial intelligence, and so on. We will find that they are special cases of the same underlying principles.