The Self-Programming System: A Skepticism-conformed Computational Framework of the Mind

Fangfang Li
Liff1229@hotmail.com
Galileo Mind Research,
Room 14-1, Unit 4, Building 13,
Taibei City, University City, Shapinba District
Chongqing, China

Xiaojie Zhang
10802@cqmpc.edu.cn
Chongqing Medical and Pharmaceutical College
Room 407, Main Building, Chongqing Medical
and Pharmaceutical College, Chongqing, China

The Self-Programming System: A Skepticism-conformed Computational Framework of the Mind

Abstract

How the mind works is the ultimate mystery for human beings. To answer this question, one of the most significant insights is Kant's argument that we can only perceive phenomena but not the essence of the external world. Following this idea, phenomenologists like Husserl advocate suspending the reliance on the existence of objective reality. By adopting this attitude, we formulated a novel computational framework to model the mind phenomenologically. Specifically, we adopt two assumptions: 1) Without assuming the existence of the external objective world. 2) Symbolic intermediate-level representation is necessary. Under these two assumptions, symbols represent the persistent coupling of relationships between senses and actions rather than external objects. Following this insight, we establish a computational framework to interpret the mind, which we call the self-programming system. We also articulate how this system can naturally generate the concepts of time, space, causality, and consciousness. Besides that, we also draw a conclusion to the mind-body problem. Specifically, viewing subjective feelings as the basic existence is compatible with that there exists an external world and the body play as the substrate of the mind. Since the relationship is "compatible" rather than "cause", skepticism can never be ruled out in principle. The self-programming system is the first symbolic and programmatically implementable framework that conforms to skepticism. Thus, it may initiate a new starting point for understanding the mechanism of the mind.

Keywords: philosophy of mind; space; time; skepticism; externalism; causality; consciousness; symbolic computing; decision-making; mind-body problem;

1. Introduction

How the mind works? The first attempt at this question is too early to be traced back. We can even conjecture that, no matter in the West or the East, as long as there was civilization, there have been thinkers who tried to give replies to this question. However, no answer is eligible enough to provide a principle for practical tasks like creating a human-like intelligent agent. Or even worse, the direction toward such a goal is still obscure.

To this question, modern researchers' approaches can be divided into four categories by the answers to the following two questions:

1) Whether assume the existence of the objective external world?[1]

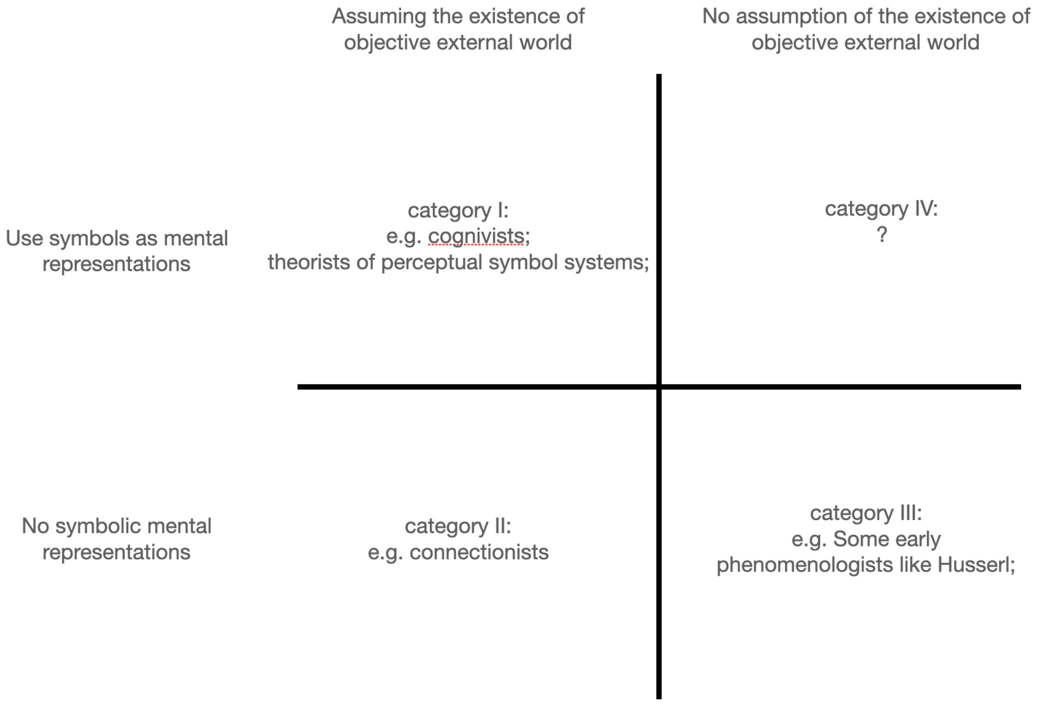
2) Do mental representations consist of mental symbols ?

Most researchers advocate for Category I or Category II, but some early phenomenologists like Husserl follow the basic tenets of Category III. We propose pursuing the proposition of Category IV, which suggests that the mind can be understood as a system that employs symbols to encompass the persistent relationships between senses and actions *without assuming the existence of an external world*. In this article, we will formulate a computational framework to understand how the mind works under these assumptions.

Since the implication of not assuming the objective external world is unclear, one good way to clarify it is by comparing it with similar ideas in Category I. Here, we choose the perceptual symbol system (Barsalou, 1999), a modal symbol system that also asserts that symbols represent senses and actions.

In the perceptual symbol system, one key argument is that the simulator that represents concept is established by repeatedly observing objects in the same category. This is to say that there has been a pre-assumed objective category. And the concept in the mind is representing it.

[1] Some readers may argue that the statement 'no assumption of the external world' is fundamentally wrong, as it suggests that we could be a 'brain-in-a-vat.' In response to this, we maintain that the brain-in-a-vat hypothesis is compatible with our theory. However, this does not mean that our theory is necessarily incorrect. In fact, we will demonstrate that the arguments presented by Putnam and his supporters, which refute the possibility of the 'brain-in-a-vat', are problematic in section 5. Moreover, our conclusion is that even if we are living in the real world and the mind is the product of this real world, we can never prove it in principle (This is analogous to the conclusion of Godel's theorems in Math). In this sense, the self-programming mind is a computational model that accords with skepticism..



	Assuming the existence of objective external world	No assumption of the existence of objective external world
Use symbols as mental representations	category I: e.g. cognivists; theorists of perceptual symbol systems;	category IV: ?
No symbolic mental representations	category II: e.g. connectionists	category III: e.g. Some early phenomenologists like Husserl;

Figure 1

In contrast, our approach does not rely on the assumptions of the existence of the external world, but instead rely on the assumption of pre-existing subjective senses and actions. Thus there must be another internal anchor for the establishment of concepts. This anchor can be loosely described as the persistent relationships between senses and actions.[2] Specifically, once these relationships have been identified, symbols representing these relationships will form.

We can see that the difference being discussed here is not just a matter of ideological perspective but also has significant implications for computational models. When symbols are thought to represent objective things, they are established based on observation of those things. However, when symbols represent persistent relationships between senses and actions, their establishment will involve in a complex and self-evolution process that depends on purpose, context, and past experiences. This is the process that we call *self-programming*. The article aims to provide a framework for capturing such a process.

Another key point to note is that we cannot equate the subjective senses with the signals transmitted by sensors. This is not feasible under the assumption of the self-programming system. Equating the

[2] Some may think the relationship between senses and actions can be expressed easily. For example, we can intuitively connect one sense to another sense through action. However, this one-to-one relationship is far from the whole picture. The more common relationships are characterized by one set of senses connected to another set of senses. For example, we need perceptual information from both vision and olfaction to distinguish a real apple from a wax apple and determine whether it is edible. This set of senses not only contains multiple senses but also has internal structures. Specifically, a person may need to first determine the location of the apple through vision and then perform the action of smelling to obtain olfactory sense. Therefore, this set of senses here actually contains an internal structure connected by actions. This implies a recursive data structure is needed to capture such relationships. So, we can only loosely describe this anchor as the persistent relationships.

senses with sensor signals is akin to assuming the existence of an objective world, and carving out the subjective aspect of senses. The result is that the theory would lose its ability to explain the phenomenal aspect of consciousness. In the setting of the self-programming system, *the senses and actions are regarded as the basic elements of cognition.*

The last implication of not assuming the existence of the objective world is that we cannot use the degree of proximity between subjective sensible properties and objective properties to define whether such senses can represent the corresponding objective properties. For example, we cannot take the sense of measurement closest to objective time within the body as our definition of time in thought.

In summary, the phrase "not assuming the existence of the external world" have three implications:

1） Since there are no objective categories, concepts cannot anchor on objective categories.

2） Senses are not equal to information transmitted by sensors. They are basic elements of the mind.

3） The degree of proximity between subjective sensible properties and objective properties is not the standard for deciding whether such senses can represent the corresponding objective properties.

If we compare traditional solutions and the idea of the self-programming mind from the viewpoint of the problem they address, we can see that they answer two different questions. The former addresses the question of 'What algorithm should an agent have to cognize the external world and work like a human?' The latter, on the other hand, addresses the question of 'What algorithm should an agent have to work like a human, including having concepts of time, space, causality and consciousness?' Since we are addressing the latter question, it is our responsibility to provide explanations of time, space, causality and consciousness, which will be presented in sections 3, 4, and 6.

Since the ‘not assuming the existence of the external world’ assumption is a fundamental but uncommon basic setting for computational accounts of the mind, this may require readers to temporarily forget the knowledge and terminology they already know under alternative assumptions for better understanding.

Although most researchers belong to Category I and II, there are also some early phenomenologists like Husserl who advocate suspending assuming the existence of the external world. They call this methodological attitude ‘bracket’. However, more recent phenomenologists and enactivists (modern cognitive scientists who agree with phenomenology, especially Merleau-Ponty’s phenomenology) partially abandoned Husserl’s advocation. For example, Merleau-Ponty believes the body cannot be bracketed. So only these early phenomenologists conform to our first assumption.

However, regarding our second question that whether mental representations consist of mental symbols, both Husserl and his later phenomenology and enactivists believe that intermediary symbolic representations are unnecessary (Gallagher and Zahavi, 2008，Gallagher，2020). As Husserl noted:

> *“..., it forgets to ask how the subject is supposed to know that the representations are in fact representations of external objects.”(P96. Zahavi, 2007)*

Or as Varela, Thompson and Rosch noted:

> *“..., symbolic computation might come to be regarded as only a narrow, highly specialized form of cognition.”(P103, Varela, Thompson and Rosch, 1991)*

These statements pointed out two key reasons why phenomenologists’ don’t believe symbolic representations play fundamental roles in cognition. The first says that there is no way to confirm a symbol indeed represents an external object that it is supposed to represent. The second says symbolic representation is not of universal benefit to cognition. Therefore, it is unreasonable to assume such a general intermediate-level representation.

As we have previously noted, symbols do not need to refer to the external world. Thus the first reason is not against the idea of the self-programming mind. To the second reason, our analysis of the self-programming system reveals that symbols function to unify the senses and the process of thought into a unique representation. This unification enable the subject to learn not only the external objects, but also her own learning mechanism. With this knowledge, the subject can enhance its procedure of learning. In other words, this unified form of expression endows humans with the ability to improve their own learning abilities through the process of learning.

In this paper, we propose a novel computational framework for understanding how the mind works based on mental symbols and without assuming the existence of the objective world. We will also explore how the concepts of time and space can be derived within this computational framework.

Furthermore, we will utilize this framework to shed light on the concept of consciousness and address the hard problem of consciousness.

Importantly, by adopting the idea of ‘not assuming the existence of the external world’, our idea offers a new perspective to understand causality. This allows for addressing mind-body problem that are typically difficult to approach with computational models.

Moreover, the self-programming system can address the unsolved symbol grounding problem, that is, how symbols acquire meaning, as proposed by Harnad (1999, also see Li and Mao, 2022). According to the idea of the self-programming system, if "meaning" refers to the representation of the external world, then this is a misguided question. Humans cannot acquire meaning in this sense. If "meaning" refers to the relationship between sense and action, then the self-programming system provides the solution.

Due to the fact that self-programming systems can determine what action to use in different situations, they are also related to decision making. However, this type of decision-making is different from the classic decision-making theory that uses probability to maximize expected payoff, as it provides a solution for decision-making under radical uncertainty. The environment addressed by this method has two key differences from that applicable to classic theory.

The first key difference is that in classic decision making, the factors that need to be considered are assumed to be clear and limited, and their acquisition is assumed to be cost-free. However, in the real-world environment that people face, the factors that may need to be considered are infinite, and determining which factors should be included is a complex process. Furthermore, every factor that is taken into consideration comes with a cost, which includes not only the cost of computation but also the cost of obtaining information.

The second key difference is that classical decision-making theory assumes that the generative model summarized from samples is stationary. In other words, the pattern of generating data in the past is assumed to be the same as in the future. However, since Hume, it has been known that this assumption is unreliable. The environment that people face is even more uncertain and volatile, making this assumption even less valid.

In recent research, many scholars have developed decision-making theories under radical uncertainty, such as the conviction narrative theory (Johnson, Bilovich, & Tuckett, 2022). The most significant difference between our self-programming theory and these theories is that the self-programming theory can explain the more fundamental principles underlying the principles described by these theories. In other words, the principles described by these theories are essentially specialized methods for decision-making under radical uncertainty learned by the self-programming system in a specific environment. And not only are these decision-making theories for handling radical uncertainty like this, but classical decision-making theory is also a specialized method that can be learned in a specific environment. Specifically, humans tend to choose classical methods when they are reliable and easily accessible, and they may choose the conviction narrative theory's method when narratives are more reliable than their own judgments and are easily accessible. There are also many other decision-making methods that people may use, even as irrational as flipping a coin to decide whether to do something. Accordingly, it is not enough to discuss these methods themselves to fully address the problem of decision-making. What is equally important is to account for why and how humans choose different methods under specific environments. In other words, a comprehensive decision making theory must be able to explain the meta-method of choosing a specific decision-making method.

In the self-programming system, no matter the method or the meta-method of decision making is part of the system's automatic operation. In other words, there is no isolated decision-making system. The problem of decision making is actually the result of separating and simplifying this automatic system. For example, classical theory is modeling through the artificial choices of the most likely relevant factors, while ignoring the observation and computation costs of these relevant factors.

Moreover, since the decision-making system is artificially defined and is essentially a facet of a larger integrated system, this implies that focusing solely on solving the decision making problem is not sufficient to achieve a decision-making theory that can fully capture humans' processes of making decisions.

2. The Primary Ideas of the Self-programming System

In this section, we will articulate how the self-programming system works. Specifically, we will divide the following content into three parts:

1) Define the components of this framework.

2) Explain the runtime procedure of the self-programming system.

3) Introduce its learning mechanism.

2.1 Basic operations and Basic senses

We first introduce the basic elements composed of Basic Operations (BOs) and Basic Senses (BSs). In the general-purpose computer, basic elements are predefined symbols in the computer's language, like logical operations, mathematical operations, numbers, identifiers, etc. But in our framework, basic elements have completely different meanings.

Specifically, both BOs and BSs refer to certain signals can be send and receive by peripherals. These peripherals can refer to a certain part of the body, or they can refer to a module in the brain, such as a module that generates emotions.

So what are the BOs and BSs that peripherals provide? Generally speaking, since the functions of each peripheral are different, the BOs and BSs provided by each peripheral are also different. For the eyes, a BO can be rotation, positioning, focusing, and so on. A BS of the eye can be certain color blocks or a specific shape. For limbs, a BO can be some kind of rotation or movement. A BE can be moving to a certain angle or some tactile signal and so on.

There are three points in this setting need to be emphasized. First, both BOs and BSs can be viewed as symbols. These symbols accompany by a look-up table to indicate signals from the most basic neural network, like shape detection, edge detection, etc. The advantage of this setting is that the form of the schemas organizing these basic symbols is independent of the specific existence of the components of the brain and body that provide these symbols. Thus, it enables functions from various sensations can be expressed uniformly. In this sense, the self-programming system indeed establishes a schema composed of symbols that can depict relationships between all sensations.

Second, applications of this schema don't need knowledge about the lookup table. One may doubt this conclusion by arguing: if you don't interpret the internal representations by virtue of the lookup table, how can you know the true phenomenon happened in the objective world? In fact, the reason for this question is that it is presupposed to seek objective truth from the perspective of a third party. But, in fact, the mind does not need such conversion, because phenomena and the relationships between these phenomena already have been expressed internally. Thus the mind can carry out various thinking activities directly through internal expressions, such as planning, judgment, etc. In this case, objective reality is not a necessary factor for the functioning of the mind. This feature further implies the robustness of the self-programming system against the disturbance of the look-up table, since changes in the look-up table will lead to corresponding modifications of the schema.

Such independence is also applicable to time and space. This means all these relationships are only based on basic elements from senses and actions. No objective time and space context are presumed in this system. This view is different from the current mainstream building of schema. Specifically, the mainstream representations of schemas are relying on the form of the existence of these components. For example, body schemas are encoded in 3D space (Morasso et al., 2015; Macaluso & Maravita, 2010).

Third, a basic element does not necessarily correspond to a unique stimulus. A particular stimulus may correspond to a set of them. For example, one BS may represent a circular area that appears on the retina, while another BS represents the size of the area on the retina. Neither of these two symbols, respectively, can identify any unique retinal stimulus. But the combination of them can correspond to this stimulus.

2.2 Storage Object, Property, Operation and the Storage system

In the next, we will first define four fundamental concepts and then make further analysis on this basis:

Storage object[3]: The intuition of the storage object is the unit to store the relationships between senses and actions. Technically, it is composed of a set of properties.

Property: Properties need to play two roles. The first is to determine whether a bunch of stimuli from the external or internal is enough to locate an existing storage object that contains these properties. The second is that, once a particular storage object is located, these properties in this storage object can predict the outcomes of placing certain operations on the origins of the stimulus that triggered this storage object. Technically, a property is composed of

1) Storage objects or BSs;

2) Operations or BOs that connect these units in 1).

In this sense, properties are both the locators and the instructional manual of an object.

[3] Actually, storage objects are just concepts in the self-programming system. Nevertheless, since concepts in the commonsense have too many other usages and ambiguous meanings, we choose this new terminology.

Operations: a sequence of other operations or BOs that can be executed under specific conditions; these specific conditions refer to properties that the storage object associates with this operation must have.

Storage system: It consists of two parts, one is a collection of all storage objects, and the other is some specific operations that can retrieve and compare information stored in this storage system.

At first glance, the above definition seems to have a circular definition problem. However, if we think in terms of construction, the above definition is logically clear. The reason is that these definitions can be built up step by step starting from basic elements. Specifically, the combination of BOs and BSs is sufficient to construct a sequence of operations and their results. Thereby, properties are constructed. And multiple properties actually form a set of conditions, which can be combined with a sequence of other BOs to form a new operation. In other words, the conditions of an operation are actually constructed gradually in order, that is, the properties constructed first become the conditions under which the new operation can be created. The same method can also be used to construct storage objects, that is, starting from a storage object only containing a single property, and gradually defining more complex storage objects. (See Figure 2)

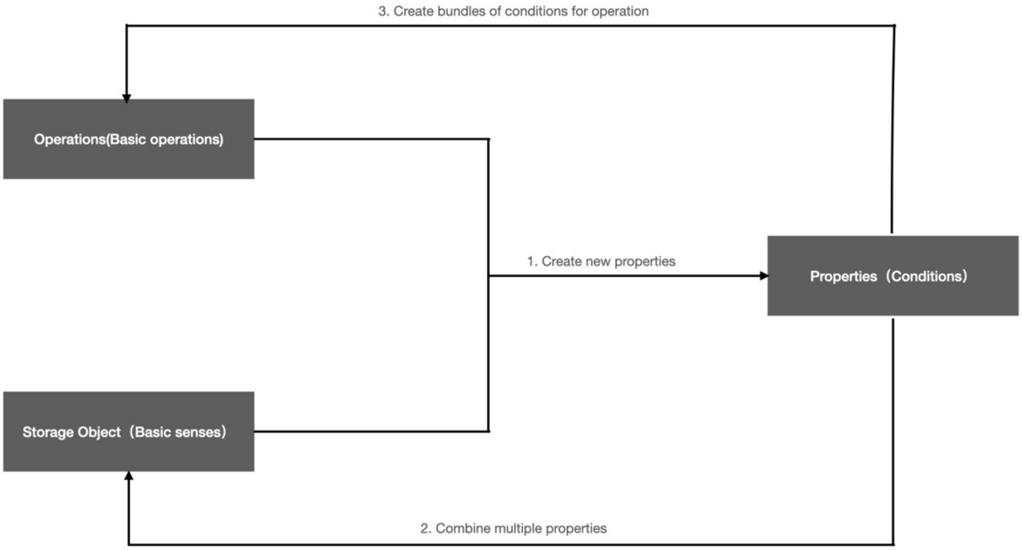

3. Create bundles of conditions for operation
Operations(Basic operations)
1. Create new properties
Properties（Conditions）
Storage Object（Basic senses）
2. Combine multiple properties

Figure 2 The relationship between operations, properties and storage object

2.3 The runtime of the self-programming system

Based on the static structure of the storage system, we can now turn to the dynamics of the self-programming system. The running of a self-programming system can be summed up in one sentence:

it is a mapping from a runtime state to an operation. We have already talked about the definition of operation, but what is the runtime state?

The runtime state is a space that can be divided into two parts, the explicit state and the implicit state. The explicit state can contain a set of active storage objects and their relations with each other that express what is currently perceived through observation, perception, feeling, thinking, etc. For example, if someone saw a plate on the table with an apple in it, his/her explicit state will include these storage objects that represent the apple, the plate, and the table, and the network that represents the positional relationship between these three. In this case, the explicit state represented the observed state of the external world. It could also represent the current internal state, for example, the current mood or the feeling, like hunger. At the same time, in the explicit state, there is also a goal. For example, when you are hungry, the goal can be to find a way to eliminate hunger.

Then what is the implicit state? Simply speaking, the implicit state is the relationship between storage objects in the explicit state and all other storage objects in the storage system. For example, let's say the current explicit state is that there is an apple on the table as described above, and the goal is to eliminate hunger. Then the implicit state may be: all storage objects that represent apples in the storage system can eliminate hunger by "eating it" (state 1); it could also be: there are some storage objects that represent apples indicate that apples can eliminate hunger, but others indicated not, such as existing a storage object representing a toy apple. (state 2).

The procedure of runtime is described in Figure 3. At first, the explicit state will be compared with the storage system. This will generate relationships between the storage objects in the explicit state and that in the storage system. These relationships will be sent to the implicit state.

Then, the implicit state will trigger some particular implicit operation. This implicit operation is for finding appropriate operations, which we call explicit operations. And the implicit operation will also determine how to use these explicit operations, such as direct execution or sending to the explicit state, etc.

For example, if the implicit operation corresponding to the implicit state happens to find that there is only one explicit operation that can achieve the goal in the explicit state (as in the case of state 1 in the previous example). Then the implicit operation can choose to run this explicit operation directly.

What if the implicit operation find not a single appropriate explicit operation? In some situations, there may exist multiple ways to achieve the goal? For example, if you want to calculate 324x99, you can directly use the general multiplication method, but you can also use 324x100-324 to calculate; Similarly, there may not exist any known operations in the storage system that can achieve the goal, for example, the goals like how a light-speed spacecraft can be built. There may also exist some way that can only achieve the goal with uncertainty, such as state 2 in the previous example.

In each of the above situations, there are further subdivisions. For example, in the case of State 2 mentioned above, the implicit operation may choose the explicit operation based on whether there are properties that can be easily collected and helpful for making further decisions. If such a property exists it can execute the explicit operation that can collect this property at first. Corresponding to State 2 of the previous case, it is possible to touch the apple first and decide whether to eat it.

In some cases, the state of the explicit operations discovered by the implicit operation can also be put into the explicit state for further calculations of what should be done. For example, if no possible solution is found, some attempts may be made by using the functions provided by other peripherals, such as a search that allows combining two operations together.

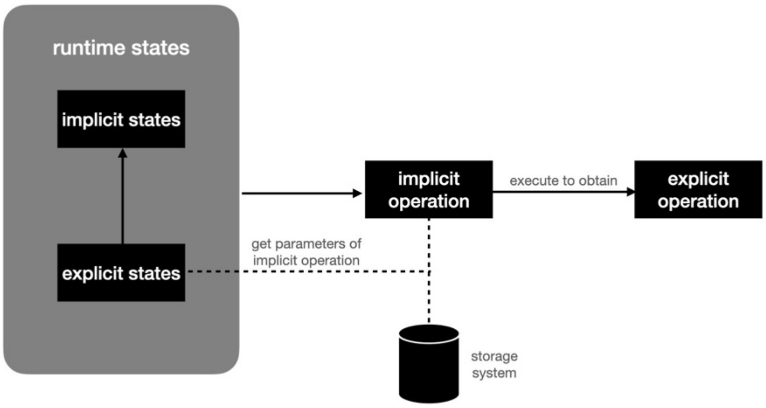


Figure 3 The procedure of runtime

In cases where there are multiple explicit operations, it is also possible to put all these explicit operations into an explicit state to determine which one is more appropriate.

To sum up, the runtime of a self-programming system provides a function that maps to the execution of specific operations based on conditions and goals. This function is obtained by comparing the current runtime state with the information in the storage system. Therefore, *the whole process of locating and executing a specific operation from the runtime state can be regarded as a Basic* *operation (BO) provided by the storage system. Since an operation in a storage system is a composition of Basic operations, this means that the operation that invokes the runtime can actually also be a possible component of the operation that compose properties.* This allows some properties of storage objects may describing how to use the storage system. This recursive structure is the most important feature of the self-programming system.

If we analogy this point to computer programming, the storage system is equivalent to providing a dynamic mapping from function names to function implementations. This dynamic mapping allows the self-programming system can set abstract goals. Then, collecting detailed information and making subtle decisions in the processing of abstract goals.

This top-down approach is consistent with how humans accomplish specific tasks. Imaging how we make a travel plan, we may first decide on the destination city and the primary way of transportation. Then, collect the prices of hotels, taxis, and others for further decisions.

Through the study of the self-programming system, we can discover some important properties. First, a self-programming system is by no means a combination of multiple domain-specified systems. The reason is that the key to realizing a self-programming system is the relationship between the storage system and external observations, and how to operate the data in the storage system under these relationships. This is a completely abstract domain that is independent of any specific domain. No matter what domain a problem belongs to, it ultimately lies in how to manipulate the data in the storage system. This means that, for any information, as long as it can be stored, it can be processed in the same way.

On the other hand, we can see that when the runtime state triggers an operation, the operation could consist of a sequence of sub-operations that may trigger new mappings. This is a process similar to fractal problems in complex science. Therefore, solving one part of a problem is no easier than the whole problem. In other words, without a proper understanding of the storage system, even trying to solve some seemingly simple problems will lead to clueless.

2.4 Learning mechanism

As can be seen from the previous analysis, if the mapping of runtime states to implicit operations and the information in the storage system are given, the run of the self-programming system will be determined. In other words, how the self-programming system works depends on the information in the storage system and the implicit mapping. There is a naturally following question that is how the storage objects and implicit mapping are established? Or what is the learning mechanism behind them?

The problem is both simple and complex. The simple part is that if the mind keeps perceiving some procedures composed of certain phenomena and operations repeating, it can distinguish these relevant phenomena and operations against irrelevant factors to form a property. Since the properties are the content of the storage object, creating properties is equivalent to creating new storage objects.

However, an answer like this can only capture a basic functional explanation of the learning mechanism. The more important question is what decides the action of perceiving since it is the one that indeed decides what storage objects to be formed. Unfortunately, facing this question, we can only answer part of it. The other part cannot be summed up by the nature of the self-programming system.

In the self-programming system, the application of any function has two different levels, namely the spontaneous level and the purposeful level. This rule is also applicable to the learning mechanism. Its spontaneous level refers to the fact that this learning mechanism is automatically triggered during the operation of the system. The role of the learning mechanism at this spontaneous level is relatively simple and can be described. It works on at least the following three aspects.

First, the most immediate aspect is to work with explicit state at runtime. Specifically, if a certain storage object happens to be triggered at some point, its properties are loaded into the explicit state. At this time, if the same result that generated by an operation happened repeatedly, then a new property that contains the new operation and the result will be created. And this new property combines with the properties from the original object to generate a new storage object.

Second, since the runtime state not only has explicit state and explicit operations, but also has corresponding the implicit state and implicit operations, the learning mechanism works should also work on the implicit aspect. That is, building mappings from the implicit state to appropriate implicit operations. Taking the previous calculation 324x99= as an example, the implicit state is that there are multiple ways to calculate this result, and the implicit operation is to list this method into the explicit state and consider it further.

The third aspect is specializing the implicit mappings. We introduce this aspect by an example. Assume there is a problem, and both operations A and B known in the system can solve it. We know that in this case both operations A and B shall be put into the explicit state to be evaluated by a more general implicit operation. Here, we further assume that the result of the evaluation is that

Operation A executes faster so Operation A is always called in more urgent situations; Operation B has a higher success rate, thus it is always called in situations with spare time. Then if these operations are called repeatedly, two new implicit mappings will be created: Calls Operation A under emergency situation. Call Operation B when there is spare time. In this way, the process of loading the implicit state into the explicit state is avoided by forming a specialized mapping, thereby reducing the computational cost.

After talking about spontaneous learning, let's turn to purposely learning. As we said before, if certain states, operations, and results occur repeatedly, then a new storage object will be generated. This newly created storage object expresses a specific function by its properties. The learning mechanism can still be viewed as a function, thus it can also be expressed by a storage object which is created by the repeat of the spontaneous learning process. The result is that a storage object that expresses the learning mechanism will exist in the storage system.

Once the above storage object is created, the self-programming system can use the learning mechanism to create new storage objects purposefully like other peripherals. In this case, the question of when to apply the learning mechanism becomes a non-summerizable question, since its application conditions are completely determined by the self-programming system itself. As we said earlier, the problem of self-programming is a fractal problem. So in this sense, summarizing it is equivalent to resummarizing the whole self-programming system.

3. The concepts of Time and Space

In traditional views, time and space are regarded as the inherent properties of the objective physical world. The concepts of time and space in the mind are merely expressions of these inherent properties.

For instance, some scholars may argue that certain systems in the biological organisms of mammals, such as those that generate Circadian rhythms, are capable of corresponding well with objective time and these systems should be regarded as the primary source of the concept of time. Similarly, with regard to the concept of space, due to the existence of well-functioning systems in the nervous system, such as the grid system, that can accurately measure objective space, these systems are considered the primary source of the concept of space.

The fundamental belief of this idea is that the degree of a system's measurement of the objective time and space determines whether the system should be considered a source of the concepts of time and space. However, if we reason based on this idea, we will encounter problems. For instance, imagine if there were another, better way of measuring objective time and space, such as implanting a mechanical clock or a GPS-like system into the body, would the source of temporal and spatial concepts undergo a fundamental change? Alternatively, we could also ask, to what degree must a system's measurement be close to objective time and space to be considered a source of the concepts of time and space?

In the self-programming system, as there is no assumption of the objective time and space, new interpretations of the concepts of time and space will be provided. Specifically, we will answer what the couplings between sense and action that give rise to the concepts of time and space are.

3.1 Time

Under the assumption of the self-programming mind, finding a substitute for objective time is easy. In fact, the concept of time is composed of *all* sequences of senses and actions that can be measured by the common sense definition of time. It should be noted that the term "the common sense definition of time" does not imply the existence of objective time. The establishment of the time concept follows a reverse procedure. Specifically, once the self-programming system detects that certain sequences of senses and actions will occur simultaneously under specific circumstances, it forms a concept. This concept happens to be called time.

The advantage of this formation is that any sequence related to the concept of time can be used as a timer. Loosely speaking, the self-programming system will choose the suitable one based on different circumstances. These sequences could be generated from some internal timer in the brain, the count of heartbeats, or even watching a clock's tick. Some of them are used to mark a long period but only require low precision, and others are used to indicate a much shorter time but need high precision. This is because some timers will be severely affected by other factors, like emotion, while others can resist these affections of time. All in all, the self-programming system will choose the best timer for different purposes and environments.

3.2 Space

Although the origin of the concept of space is absolutely different from that of time, by following the principle of coupling senses, we can also naturally speculate that the concept of space is a representation of the coupling of senses under transformation, such as translation or rotation. （see figure 4）

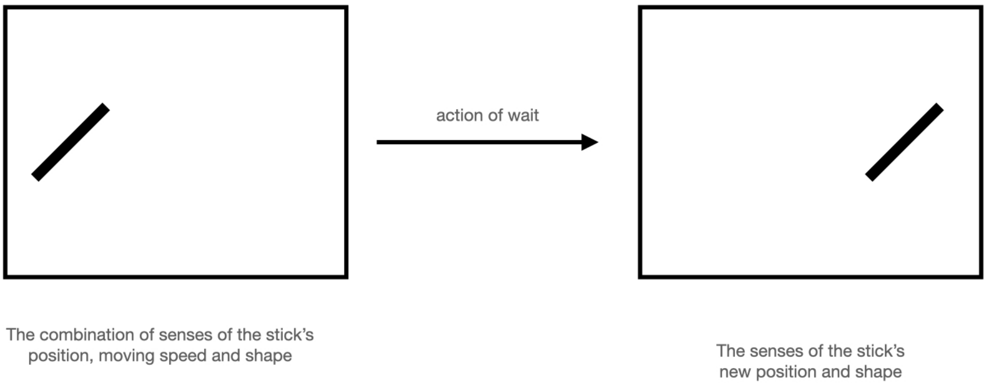

action of wait
The combination of senses of the stick's position, moving speed and shape
The senses of the stick's new position and shape

Figure 4

However, more importantly, space may also represent a linear relationship between the high-level comparison of senses. To see this, let's imagine a robot that is designed based on the principle of the self-programming mind. One day, it records what a particular tree looks like at 100 meters distance. (This distance can be described with the language of subjective senses. Specifically, it could be moving with a fixed effort with a sense of time, e.g. this effort is equivalent to the objective velocity of 1 meter per second; this sense of time is equivalent to the 100 seconds of objective time.) And it also recorded what this tree was like at a distance of 50 meters. If the perceptions of this tree at different distances have no relation, the robot cannot predict what it will look like at other distances, like 20 meters or 60 meters. But if the robot can find out that these two perceptions only differed in size. Then it can figure out there are linear relationships between the size and the distance.

The procedure of building such linear relationships can naturally be captured in the self-programming system. This is because these comparisons of sizes and distances are just the relation between two senses. Thus they will manifest in the implicit state. This further implies the implicit operation can calculate what the sense of the tree will be like at other distances.

Following the idea of the self-programming mind that a symbol represents the relationship between senses and actions, space is the symbol that represents both the coupling of senses from transformation and visual linear couplings, like the linear relationship between sizes and distances.

4. Causality

The 18th-century British philosopher David Hume questioned whether the law of causality was real. In Hume's view, causality cannot be directly observed, so he believes that causality is just an illusion in the mind. Take the following example "the branch of the apple tree breaks and the apple falls". these two events that we can observe are "the branch breaks" and "the apple falls”. But whether there is a causal relationship between the two events is unobservable. From this, Hume further concluded that causality is nothing but a name for the empirical correlation that is always adjacent in time and space. This view of causality proposed by Hume is called "the constant conjunction theory”.

But can this theory explain causality? The answer is no. For example, crows of roosters always occur when the sun is about to rise, but we obviously don't think of roosters as the reason for the sun's rise. This shows that the events that happened one after the other in time may not have a causal relationship.

To look at another example, if you find your alarm doesn't go off one morning, it's most likely because you forgot to set your alarm last night. In this case, there is no temporal adjacency between these two events “the alarm didn't go off” and “forgot to set the alarm”.

From the above two examples, we can see that Hume's constant conjunction theory is neither a sufficient nor a necessary condition of causality. In fact, Hume himself was not unaware of the problems of this theory. He once noted:

> *“We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or, in other words, where, if the first object had not been, the second never had existed.”*

However, later researchers pointed out that Hume's statement actually pointed to two different theories, not simply "in other words”. Because the second half of its statement —if the first object had not been, the second never had existed— points out a whole new definition of causality. This definition is called counterfactual theory by later researchers (Pinker, 2007).

While scholars generally agree that counterfactual theory is a more tenable theory than constant conjunction theory, critics quickly pointed out that there are fundamental problems with the counterfactual theory. Critics argue that we cannot replay our lives like movie replays, so it is impossible for us to observe another situation in a world that did not happen. So, for a counterfactual theory to hold, the key task is to explain what this unobserved possible world is.

This task can easily be done in the self-programming system. Specifically, this unobservable possible world is exactly the storage object. Recall a property in a storage object represent what will be the outcome if a particular action are placed on the object. Since a storage object contain multiply properties, every property indict a different but possible outcome. And all these outcome are unobserved, since it has not happen but it could happen. Accordingly, the storage object is the unobserved possible world in the counterfactual theory.

We can adopt an example to examine this idea. Take a storage object representing a vase on a table. It contains Property A which is "hit it with a hammer and it will break" and another Property B which is "leave it there and it will stay fine”. If the event represented by Property A occurred, then there is a counterfactual contrast to the event represented by Property B that did not occur. Thus the behavior "hitting with a hammer" in Property A becomes the reason for the broken vase. To sum up, the unobservable possible world of counterfactual theory actually exists, but it exists in the storage structure of the mind, not in the real world.

Based on the analysis above, we can conclude that *causality is essentially a way of organizing the relationship between senses*. An important implication of this conclusion is that *if an object is theoretically imperceptible, any causal reasoning based on it would be invalid*. For instance, Kant's concept of Ding an Sich falls in this category. According to Kant, Ding an Sich is the cause of perceptions. Under this definition, Kant separates the sensible features from the Ding an Sich, which implies that Ding an Sich itself is not sensible. However, Ding an Sich is also seen as the cause of perception. Therefore, in this case, the inference of the existence of Ding an sich is invalid. Another example is Merleau-Ponty’s advocation that phenomenology cannot bracket all external existence[4]. Specifically, he emphasizes that there must cause play as the substrate for the perception. This cause is the subjective aspect of the body. So the body cannot be bracketed. In this argument, we can see that the subjective aspect of the body is not perceptible principally (otherwise it is the objective aspect of the body). Thus the subjective side of the body is actually the same as Kant’s Ding an Sich. This further implies that the causal inference of the existence of the subjective side of the body is invalid.

In addition to theories like counterfactual theory, there are also other theories that attempt to explain causality from a probabilistic perspective, such as causal modeling (Pearl, 2010). However, using probability models to explain fundamental problems related to human thinking inevitably involves the problem of how to deal with radical uncertainty environment, which we have previously introduced when discussing classical decision making theory. This further implies that, like classical

[4] In Husserl’s phenomenology, Bracket the external world means stop assuming the existence ot the external world.

decision making theory, causal modeling can only interpret causality in specific situations rather than providing a comprehensive explanation.

5. Skepticism, Externalism and Brain-in-vat

Skepticism has traditionally been regarded as a philosophy of negation and has thus been a target of refutation by mainstream philosophers. Philosophers have attempted various methods to evade skepticism. However, this paper shows that skepticism can also provide insights for building computational model and hence can be constructive. Additionally, criticisms of skepticism can also be analyzed and its loopholes can be uncovered from the perspective of our computational models.

One particularly vivid refutation of skepticism comes from Putnam's famous thought experiment, "Brain in a Vat" (BIV). Putnam sought to demonstrate the impossibility of the Brain in a Vat scenario, thereby proving the fallacy of skepticism.

The setting of the BIV thought experiment can be described as follows:

> *Consider the hypothesis that you are a disembodied brain floating in a vat of nutrient fluids. This brain is connected to a supercomputer whose program produces electrical impulses that stimulate the brain in just the way that normal brains are stimulated as a result of perceiving external objects in the normal way. (The movie 'The Matrix' depicts embodied brains which are so stimulated, while their bodies float in a vats.) If you are a brain in a vat, then you have experiences that are qualitatively indistinguishable from those of a normal perceiver. If you come to believe, on the basis of your computer-induced experiences, that you are looking at a tree, then you are sadly mistaken. (McKinsey, 2009, Skeptical Hypotheses and the Skeptical Argument, Para. 1)*

Putnam concludes that BIV is impossible. His main argument is that the meaning of words is external, which means it partially depends on what external objects are. Therefore, when the Brain in the Vat uses the term "I see a tree", the meaning of this statement is different from the same statement said by a person living in the real world. The Brain in a Vat's representation of the tree is only a result of the stimuli provided by the supercomputer, and not the actual tree itself. Thus, since what we see as a tree is indeed a tree, we cannot be a Brain in a Vat.

Indeed, one crucial reliance of Putnam's argument is that the meaning of words must be external, which means that the meaning of words is inherently related to the external world's objects. This idea that there is a necessary relationship between the meaning of words and the external object is known as semantic externalism. Therefore, to demonstrate the validity of Putnam's refutation to BIV, one must prove that semantic externalism is reasonable. Thus, Putnam promoted another thought experiment to validate semantic externalism:

> *Imagine a planet, a twin earth. There are a few things you should know about this planet. First, it duplicates earth in almost every respect. ... The twin version of you is your exact, molecule-for-molecule, duplicate. Second, the only difference between earth and twin earth is this: on twin earth, the substance that sits in oceans, flows in rivers, and comes out of faucets is not water. It is not water, because it is not made up of two parts of hydrogen to one part of oxygen. In fact, it is not made of hydrogen and oxygen at all but, rather, some other elements entirely absent on our earth. Nevertheless, third, this substance looks, tastes, and feels exactly like water does to people on our earth, and is used in the same way. Finally, this thought experiment is set in 1750 (and, a corresponding twin 1750), before people had any idea of the underlying molecular structure of substances such as water. (Rowlands, Lau, and Deutsch, 2020, Arguments for Content Externalism, Para. 2)*

Putnam argues that even though people on Earth and on Twin Earth use the word "water" to refer to water and twater (the water-like liquid on Twin Earth), they have different meanings. This is because their essence, namely their chemical composition, is different. He, therefore, concludes that the meaning of words is external, as it is related to external factors such as the chemical composition of water.

However, one major flaw in Putnam's argument arises when we consider both of his thought experiments together. The BIV and the Twin Earth thought experiment are interdependent with each other. Specifically, to refute the BIV, one needs to demonstrate the existence of the external world, which relies on semantic externalism. However, in the Twin Earth thought experiment, semantic externalism presupposed the external world. This creates a circular argument.

While Putnam's argument for externalism suffers from the problem of circular reasoning, it does not necessarily mean that semantic externalism is false. Even from the most intuitive standpoint, it is difficult to refute the idea that the meaning of the word "tree" must necessarily have some connection to the trees in reality. The mere existence of this connection should be sufficient evidence for externalism. Putnam calls it the causal connection. Indeed, further research by scholars has found that the causal connection is the most crucial reason for refuting skepticism. As long as the causal connection is correct, skepticism is seriously problematic (McKinsey, 2009, conclusion, para. 3).

However, does this causal connection hold up? Based on our previous discussion of the nature of causation, it is not valid. The "real" tree that Putnam refers to is actually an unperceivable assumption (if it were perceivable, it would have already been internalized). This assumption is essentially the same as Kant's King an sich. Therefore, this causal connection is meaningless.

Furthermore, we can also reinterpret the twin earth thought experiment from the perspective of self-programming theory. The self-programming theory suggests that 'water' and 'twater' have the same meaning when their chemical structures are not distinguishable. This is because the meaning of 'water' is defined by the structure that is a composite of basic senses and operations. Since both of them have been precepted as the same, these two structures are exactly the same. However, once someone invents and uses a device that can distinguish them chemically, water and twater will have two different properties respectively. Both these two properties indicate the way of using this device but with two different outcomes. One outcome is the H2O and the other indicate the chemical structure of twater. Then, for the experimenter, the concept of water will be split into two concepts. This is actually a typical process of creating new concepts that have been included in the mechanism of the self-programming mind.

In addition to Putnam's refutation of BIV, Daniel Dennett proposed another way to refute it (1991). In simple terms, Dennett believed that simulating a real world for BIV is too complex. For example, picking up a handful of sand and letting it slip through your fingers. The physical relationships between the sand particles are very complex due to their mutual interactions, requiring a huge amount of computation. Therefore, it is impossible for a computer to simulate the real world.

The easiest way to echo Dennett's view is: the time for the brain in a vat does not necessarily be the same as the time of the computer simulating the world. Even if such a simulation takes longer than the entire age of the universe, there would be no problem. This is because the universe itself could also be virtual.

However, the fundamental problem with Dennett's reasoning is that it is based on the assumption that our perception is the result of material processes that cannot be perceived. This type of reasoning, like Kant's thing-in-itself, goes beyond the limits of causal inference.

It is important to note that the argument we are trying to express in this chapter is not that we are in fact brains in vats, but rather that we cannot prove that we are not. More importantly, if there is a real external world from which creatures that act like self-programming systems are born, they also cannot prove that they are living in the real world. In other words, the unprovability and the existence of an external world are compatible. Based on this reasoning, we can regard the fact that we, humans, are creatures of the external world we observe as a self-evident axiom. With this axiom, all empirical sciences become meaningful.

6. Consciousness

What is the nature of consciousness? This question, like how the mind works, has haunted all intellectuals since ancient history. In this section, we will first answer this question by employing the self-programming system, then solve the well-known hard problem of consciousness by showing why we cannot figure out subjective feelings from an objective perspective.

6.1 The Nature of Consciousness

Why does consciousness so hard to be interpreted? The reason is still rooted in the common misunderstanding of symbols since consciousness is also a symbol in the mind.[5] In fact, if we treat external objects as the basis of cognition, no consensus can be reached on this problem. Researchers' argument can be divided into the following four categories.

The first category holds the view that there is no subjective conscious experience (Rey, 1986; Dennett, 1991). However, this view is inconsistent with our experience.

The second class of view is that there exists conscious experience and it can be explained objectively. (Churchland, 1986; Crick, 1994; Koch, 2004; Hurley, 1998; Noë, 2005, 2009). The main problem with such a view is that they fail to explain that we seem capable of producing a mechanism with the same function but without consciousness.

Research in the third category acknowledges that conscious experience exists and it is not scientifically explainable. However, they believe such inexplicability is not so significant. We only need to focus on how to connect consciousness experience to physical stimuli (Block, 2002; Block

[5] Some scholars may argue that consciousness is not a symbol, but a process. However, the statement that consciousness is a process has no practical implication because any events of the body, whether conscious or unconscious, can be considered as a process. To regard it as a symbol implies that it consists of persistent couplings of senses and actions. Specifically, in the self-programming system, the sequence of triggered storage objects (composed of senses and actions) in the explicit state will be recorded. And this recorded sequence can be sequentially traced by an internal action. Since such trace action leads to a fixed sequence of triggered storage objects. Thus it is a persistent coupling of the senses of the triggered storage objects and the action of tracing. Thus it constitutes a symbol.

and Stalnaker, 1999; Hill, 1997; Loar, 1997, 1999; Papineau, 1993, 2002; Perry, 2001). The biggest weakness of this interpretation is why the consciousness is as unusual as inexplicable.

The fourth category is dualism, that is, the world has both physical and consciousness. So it is not surprising that consciousness cannot be explained physically. This view can be traced back to Descartes. But this view is generally not accepted because it is divergent from the current scientific paradigm (Collins, 2011). Another alternative view is that although there are both physical and phenomenal objects, phenomenal experience does not have an impact on the physical world (Campbell, 1970; Jackson, 1982; Robinson, 2004). The natural question of this viewpoint is why there is such a non-necessary phenomenal experience.

However, if we transfer our standing point from objective-existence-based cognition to sensorimotor-based cognition, the nature of consciousness can be understood clearly. Next, let's analyze it from this perspective.

As we noted at the beginning of this article, symbols represent the relationships between sensorimotor. Then when we introduce how the self-programming system works, we regard these operations in the thinking process as the same as the bodies' operations. Consciousness is undoubtedly a symbol. Thus it must a representation of relationships between these Basic operations and Basic senses. The problem is just what these operations and elements exactly are.

Here, we adopt a usual definition of consciousness, which is the ability of a subject can experience objects. Since we have assumed any symbol represents couplings of senses and actions and symbols are the origins of objects, the ability to experience objects is just experiencing a bundle of senses. Since senses are by definition something for experiencing. Thus experiencing objects is not a special ability. What really distinguishes "the conscious" and "the unconscious" is whether the subject knows these senses have been triggered. In other words, the distinction is whether these triggered senses have been recorded for retrospection in the future. This will lead to the question -- what bundle of senses will be recorded?

Our answer is all storage objects have been put into the explicit state will be recorded. This conclusion can be validate both functionally and empirically.

From the functional perspective, the intention of putting a storage object into the explicit state space is to explore its relationships with other storage objects in the storage system. And using these relationships to locate and run a particular implicit operation. Such operations usually need to be placed on the storage object that triggered this implicit operation. This means that if the storage object in the explicit state is not recorded, this particular implicit operation cannot locate the target storage object. This will lead to the failure of these operations.

From the empirical evidence, various existing neuroscience-based theories about the functionality of consciousness are consistent with our ideas. (Seth and Bayne, 2022) Among these theories, Global Workspace Theory (GWT) is the most influential. It regards consciousness as a global space for information interaction. (Baars, 1988, 1997, 2002; Dehaene & Changeux, 2011; Mashour, Roelfsema, Changeux & Dehaene, 2020) The information in it will be broadcast to various subsystems, thus these subsystems can be combined to determine the optimal behavior globally.

Another influential theory is the higher-order theory (HOT). The core idea of these theories is that if some information is conscious, then it must be the information for meta-representation. (Brown, Lau, & LeDoux, 2019; Rosenthal, 2005) The meta-representation here refers to a description that is not a direct description of the world but a higher-level description that goes beyond objective facts. For example, "yesterday, the vase was broken and seriously affected my mood." In this case, the broken vase is a description of the objective world, and the whole sentence is a meta-representation beyond the objective.

In the self-programming system, storage objects in the explicit state space are for comparison with other storage objects for abstracting relationships. Such relationships are exactly meta-information. Thus our conclusion is consistent with the idea of HOTs.

And, since the storage system possesses all knowledge that the subject knows, an operation triggered by the comparison with the current environment and the storage system has already been considered in the global scope. This point is also consistent with GWTs.

In summary, we conclude that the nature of consciousness is just the action of putting storage objects into the explicit state space.

6.2 The hard problem of consciousness

Based on our previous conclusion of the nature of consciousness, we can now discuss the well-known "hard problem of consciousness". (Chamlers, 1996; Nagel, 1974; Levine, 1983, 1993, 2001) It asks why there seem to exist objectively inexplicable feelings of consciousness. We will see that this is just a matter of course based on the idea of the self-programming mind.

Let’s begin with defining several required concepts:

1) What is objective?

2) What is explanation?

To define “objective”, we need to define “self” first. In fact, we already discussed in the learning mechanism section that the reason a storage object is formed is to pack the properties of the object being perceived. Thus a storage object expresses the observed object. If the observed object is a body part, then there will be a storage object representing the body part; if the observed object is an external being, then there will be a storage object expressing the external being. So what if the object being observed is the self-programming system itself? Then the storage object formed will express all the content that appears continuously in the explicit state. Since we already know the content of the explicit state is actually a result of both implicit manipulation and external stimuli based on the body. Therefore, this storage object can represent a subject's whole experience of the mind. Thus, it expresses the subjective self.

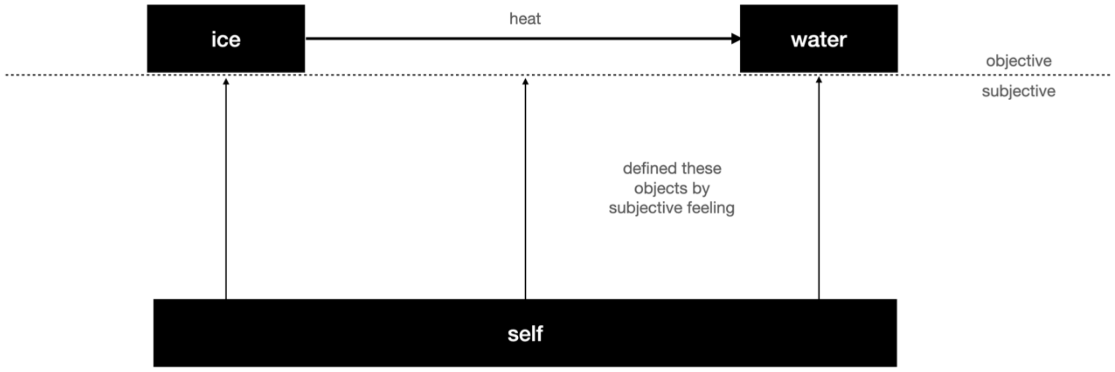


Figure 5: Take the melting of ice as an example. In this example, the objective process only includes the heat of ice, then result in water. However, ice, water, and the process of heat are defined based on the feelings of the "self", whereas the "self" does not belong to the objective world.

Combining this conclusion, we have already known that both the representations of the external world and the self are storage objects in the storage system. And they are connected by properties with each other. Since we also know the commonsense of “objective” is something irrelevant to the subject, we can naturally infer "objective" represent the remaining part after all the properties connected to the self are removed.(See figure 5)

Then, let's look at the nature of interpretation. The so-called interpretation is actually that some observed properties can be deduced from other properties. These properties that are used to deduce are called basic laws. Because the objective representation of the world is what remains after removing properties associated with the self. Therefore, the basic laws of the so-called objective interpretation must be properties in the part that has no property related to the storage object of self.

However, we also know that the self is the collection of all subjective experiences. Therefore, any basic laws that can explain subjective experience necessarily require the inclusion of the subjective experience of the basic elements of the cognitive system which must be related to the self, so they cannot be contained in the basic laws of the objective part. This means that objective laws cannot be used to explain subjective experience. So, from the perspective of the self-programming mind, the inexplicability of consciousness by objective analysis is the inevitable result of the nature of consciousness.

7. Empirical evidence

The proposition of self-programming systems is essentially a framework for how thought can be computed. If researchers hope to provide compelling direct evidence to demonstrate its correctness, they must delve into the entire workings of the brain based on an understanding of the meanings represented by various neural structures. Clearly, such forms of verification are still far from being attainable given current technology and understanding of the brain. Even if we were to settle for less and only verify some key hypotheses of the theory, such as whether a structure representing the overall characteristics of the storage system can be found in the brain, this still requires a deeper understanding of how concepts are represented in the brain. However, even this is currently beyond our technological capabilities.

Although direct evidence cannot be provided, there is no shortage of indirect evidence, some of which even comes directly from our lived experience. Firstly, part of the basic settings of self-programming systems bears resemblance to certain theories, such as the perceptual symbol system, which holds that symbols are used to represent senses. Therefore, empirical evidence supporting the perceptual symbol system in this aspect may also support the self-programming system. For example, a recent study on the neural representations of concepts (Fernandino et al., 2022) provides evidence in support of both the perceptual symbol system and the self-programming system.

The most significant difference between the perceptual symbol system and the self-programming system - how concepts are formed - is actually supported by more direct evidence. In fact, we do not even need to conduct experiments; starting from our own personal experiences, we can observe that learning new concepts is influenced by our prior knowledge of other concepts, and the more similar a new concept is to ones we already know, the easier it is for us to grasp. This observation is inconsistent with the notion that concept formation is simply the result of repeated observation. Instead, as the self-programming system posits, learning different concepts involves different processes. Once these processes are established, learning similar concepts becomes easier.

In fact, the phenomenon of learning that makes future learning easier has long been noticed in artificial intelligence research. It is considered a human ability that is not yet present in current AI systems, and this ability is referred to as "self-improving" or "learn to learn". (Hall, 2007; Schmidhuber, 2003)

8. Future work to be done

In addition to the work on empirical validation discussed in the previous section, the computational framework of the self-programming mind actually provides a novel perspective of understanding cognition. Therefore, some new conclusions may be drawn out by applying this theory to various domains of cognition, such as attention, working memory, long-term memory, language, problem-solving and etc.

On the other hand, there is also a great deal of work that needs to be done on the self-programming system itself. As a system for automatically organizing existing senses and actions, the self-programming system takes senses and actions as presets. However, it remains unclear what specific senses and actions are included. Some senses and actions are explicit, such as those derived from the body, but there are also implicit senses and actions present in the brain that are not obvious. Although these implicit senses and actions are likely already being used spontaneously by humans, we do not possess the ability to directly traverse all of them. Therefore, exploring and validating these basic elements is an important task for refining the self-programming system.

Furthermore, since the self-programming system is self-accumulating, there must be an innate built-in boot program in the mind. Exploring and validating this program is also an important task in researching the self-programming mind.

Reference

1. Baars, B. J. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press.
2. Baars, B. J. (1997). In the theatre of consciousness: Global workspace theory, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies, 4*(4), 292–309.
3. Baars, B. J. (2002). The conscious access hypothesis: Origins and recent evidence. *Trends in Cognitive Sciences, 6*(1), 47–52. https://doi.org/10.1016/S1364-6613(00)01819-2
4. Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149
5. Block, Ned (2002). The Harder Problem of Consciousness. *Journal of Philosophy* 99 (8):391.
6. Block, Ned & Stalnaker, Robert (1999). Conceptual analysis, dualism, and the explanatory gap. *Philosophical Review* 108 (1):1-46.
7. Brown, R., Lau, H., & LeDoux, J. E. (2019). Understanding the higher-order approach to consciousness. *Trends in Cognitive Sciences, 23*(9), 754–768. https://doi.org/10.1016/j.tics.2019.06.009
8. Campbell, Karlyn K. (1970). *Body and Mind*. Doubleday.
9. Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory.* Oxford University Press.
10. Churchland, Patricia Smith (1986). *Neurophilosophy: Toward A Unified Science of the Mind-Brain*. MIT Press.
11. Collins, Robin (2011). The Energy of the Soul. In Mark C. Baker & Stewart Goetz (eds.), *The Soul Hypothesis: Investigations Into the Existence of the Soul*. Continuum Press. pp. 123-133.
12. Crick, Francis (1994). *The Astonishing Hypothesis: The Scientific Search for the Soul*. Scribners.
13. Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, *70*(2), 200–227. https://doi.org/10.1016/j.neuron.2011.03.018
14. Dennett, Daniel C. (1991). *Consciousness Explained*. Penguin Books.
15. Fernandino, L., Tong, J. Q., Conant, L. L., Humphries, C. J., & Binder, J. R. (2022). Decoding the information structure underlying the neural representation of concepts. *Proceedings of the National Academy of Sciences of the United States of*

America, *119*(6), e2108091119. https://doi.org/10.1073/pnas.2108091119

16. Gallagher, Shaun (2020). Action and Interaction. Oxford University Press.
17. Gallagher, Shaun & Zahavi, Dan (2008). *The Phenomenological Mind*. Routledge.
18. Hall, John Storrs (2007). Self-improving AI: an Analysis. Minds and Machines 17 (3):249-259
19. Harnad, S. (1990) *The Symbol Grounding Problem. Physica D 42: 335-346*
20. Hill, Christopher S. (1997). Imaginability, conceivability, possibility and the mind-body problem. *Philosophical Studies* 87 (1):61-85.
21. Hurley, Susan L. (1998). *Consciousness in Action*. Harvard University Press.
22. Jackson, Frank (1982). Epiphenomenal qualia. *Philosophical Quarterly* 32 (April):127-136.
23. Johnson, S., Bilovich, A., & Tuckett, D. (2022). Conviction Narrative Theory: A Theory of Choice Under Radical Uncertainty. *Behavioral and Brain Sciences,* 1-47. doi:10.1017/S0140525X22001157
24. Koch, Christof (2004). *The Quest for Consciousness a Neurobiological Approach*. Roberts & Co.
25. Levine, Joseph (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly* 64 (October):354-61.
26. Levine, Joseph (1993). On Leaving Out What It's Like. In Martin Davies & Glyn W. Humphreys (eds.), *Consciousness: Psychological an Philosophical Essays*. MIT Press. pp. 543--557.
27. Levine, Joseph (2001). *Purple Haze: The Puzzle of Consciousness*. Oxford University Press USA.
28. Li, J., & Mao, H. (2022). The Difficulties in Symbol Grounding Problem and the Direction for Solving It. *Philosophies*, *7*(5), 108. MDPI AG. Retrieved from http://dx.doi.org/10.3390/philosophies7050108
29. Loar, Brian (1997). Phenomenal states II. In Ned Block, Owen Flanagan & Güven Güzeldere (eds.), *The Nature of Consciousness: Philosophical Debates*. MIT Press.
30. Loar, Brian (1999). David Chalmers's The Conscious Mind. *Philosophy and Phenomenological Research* 59 (2):465 - 472.
31. Macaluso, E., & Maravita, A. (2010). The representation of space near the body through touch and vision. *Neuropsychologia, 48*(3), 782–795. https://doi.org/10.1016/j.neuropsychologia.2009.10.010
32. Mashour, G. A., Roelfsema, P., Changeux, J. P., & Dehaene, S. (2020). Conscious Processing and the Global Neuronal Workspace Hypothesis. *Neuron*, *105*(5), 776–798. https://doi.org/10.1016/j.neuron.2020.01.026
33. McKinsey, M. (2009) "Skepticism and Content Externalism", *The Stanford*

Encyclopedia of Philosophy (Summer 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2018/entries/skepticism-content-externalism/>.

34. Morasso, P., Casadio, M., Mohan, V., Rea, F., & Zenzeri, J. (2015). Revisiting the body-schema concept in the context of whole-body postural-focal dynamics. *Frontiers in human neuroscience*, *9*, 83. https://doi.org/10.3389/fnhum.2015.00083
35. Nagel, Thomas (1974). What is it like to be a bat? *Philosophical Review* 83 (October):435-50.
36. Noë, Alva (2005). *Action in Perception*. MIT Press.
37. Noë, Alva (2009). *Out of Our Heads: Why You Are Not Your Brain, and Other Lessons From the Biology of Consciousness*. Hill & Wang.
38. Papineau, David (1993). Physicalism, consciousness and the antipathetic fallacy. *Australasian Journal of Philosophy* 71 (2):169-83.
39. Papineau, David (2002). *Thinking About Consciousness*. Oxford University Press UK.
40. Perry, John (2001). *Knowledge, Possibility, and Consciousness*. MIT Press.
41. Pearl J. (2010). An introduction to causal inference. *The international journal of biostatistics*, *6*(2), 7. https://doi.org/10.2202/1557-4679.1203
42. Pinker, S. (2007). *The stuff of thought: Language as a window into human nature.* Viking.
43. Rey, Georges (1986). A question about consciousness. In Herbert R. Otto & James A. Tuedio (eds.), *Perspectives on Mind*. Kluwer Academic Publishers.
44. Robinson, William S. (2004). *Understanding Phenomenal Consciousness*. Cambridge University Press.
45. Rosenthal, David M. (2005). *Consciousness and Mind*. Oxford University Press UK.
46. Rowlands, Mark, Joe Lau, and Max Deutsch.(2020) "Externalism About the Mind", *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2020/entries/content-externalism/>.
47. Schmidhuber, J. (2003). *Goedel Machines: Self-Referential Universal Problem Solvers Making Provably Optimal Self-Improvements*. doi:10.48550/ARXIV.CS/0309048
48. Seth, A.K., Bayne, T. (2022) Theories of consciousness. *Nat Rev Neurosci* **23,** 439–452. https://doi.org/10.1038/s41583-022-00587-4
49. Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge, Mass: MIT Press.
50. Zahavi, Dan (2017). Husserl's Legacy: Phenomenology, Metaphysics, and Transcendental Philosophy. Oxford University Press.