American Studies XX, No. 4, (December 1990), 39-77.

FUNCTIONALISM AND ITS DIFFICULTIES

*Yu-Houng Houng**

Three major problems in the traditional philosophy of the mind are 1) what are mental entities? 2) are there other minds? How do we know the existence of other minds? and 3) what is the nature of mental processes? I shall not concern myself with the problem of other minds. The goal of this project is to investigate the modern mechanistic models of the mind, and to see if they succeed or fail in modeling the mind adequately. No comprehensive review of the traditional mind-body problem will be provided here. Several different versions of functionalism will be examined and their corresponding models of the mind will be criticized. The implications of connectionist models will be adverted to helping to build a better theoretical framework for modeling the mind.

Narrowly speaking, functionalism in philosophy of mind is a computer-analogy based theory of the mind. The major thesis of this article is that this form of functionalism is doomed to failure. On the other hand, if functionalism is understood as a form of the causal theory of the mind, i.e., treating the mind as a device that takes environmental inputs and generates behavioral outputs, then functionalism becomes a correct but very general conception of the mind. Functionalism in this form seems to be not very interesting unless further details are fleshed out. The functionalism discussed in this paper is functionalism in the first sense.

The core idea of the computer-analogy based functionalism is that the mind is a computing machine, but is the mind

---

*Assistant Research Fellow, Institute of American Culture, Academia Sinica.

really a computing machine? This question cannot be answered unless we define precisely what a machine is. If we take a machine simply as some kind of mechanical device, e.g., a children's toy music box, then it is obviously false that minds are machines. On the other hand, if 'machine' refers to a sort of biological robot which is physiologically and functionally equivalent to the mind, then it is trivially true that minds are machines.[1] Thus in attacking the so-called mind-machine problem, the word 'machine' can not be taken to refer to any artifact whatsoever. Thanks to Alonzo Church, Alan Turing, Emil L. Post, A. A. Markov, and other contemporary logicians, we now confidently believe in the existence of a "natural" set of mathematically well-defined operations which are supposed to be the exclusive set of operations computable by any existing or theoretical computing machines.

It will be made clear later that this sort of mathematically defined concept of the machine is too general to serve as a useful concept to evaluate models of the mind. It will be argued that not only the mathematical properties, but also the architectures of the machine should be taken into consideration in determining the degree of success of the models of the mind.

1. WHY FUNCTIONALISM?

Hilary Putnam is the first philosopher who puts forward the functionalistic account of the mind. In his 1960 paper, "Minds and Machines," he proposes the following thesis:

TURING MACHINE FUNCTIONALISM: The functional organi-

---

[1] John Searle explicitly holds this view. See his *Minds, Brains and Science*, Cambridge, Mass.: Harvard Univ. Press (1984). But Leibniz obviously rejects the idea that artificial automata, no matter how close it is to the natural automata, can be of the same kind as the mind. His argument is that the smallest parts of natural machines are themselves still machines, while the smallest parts of artificial machines are not machines at all. Cf. Loemker (ed.), *Leibniz: Philosophical Papers and Letters*, Syntheses Historical Library (1976).

> zation of the human mind or Turing machine can be described in terms of the sequences of mental or logical states respectively.[2]

Another important thesis in Putnam (1960) is the following:

> THE AUTONOMY OF THE MENTAL: the functional organization of the mind can be stated in mentalistic terms without reference to the nature of the physical realization of the mind.

Some explanations are in order. A Turing machine consists of a tape with unlimited length, a scanning and writing head, a finite set of symbols, and a *finite* machine table. The infinite tape which is divided into single squares serves as the working and storing memory for the operations. The scanning and writing head can move along the tape in either direction. The head can read a symbol from the tape, erase a symbol on the tape, print a symbol on the tape, move to the right square, move to the left square, or do nothing. The set of symbols, $\{S_i \mid i \in \text{natural numbers}\}$, usually consists of 1, and B (Blank). The machine table consists of a *finite* set of quadruples. Each quadruple takes one of the following forms:

$$(q_i\ S_j\ S_k\ q_l)$$
$$(q_i\ S_j\ R\ q_l)$$
$$(q_i\ S_j\ L\ q_l).$$

The first quadruple says that if the machine is in state q_i, and S_j is the currently scanned symbol, then erase S_j, print S_k on the same square, and go to the state q_l. The second says that if the machine is in state q_i, and S_j is scanned, then move one square to the right and go to the state q_l. The third says exactly the same thing as the second except that the head moves to the left instead of to the right.[3] A Turing machine then can be

---

[2] Putnam (1960), p. 373.

[3] Cf. Martin Davis (1973). The above and the subsequent descriptions of Turing

defined as a finite (nonempty) set of quadruples that contains no two quadruples whose first two symbols are the same.

We define an *instantaneous description* α_i of a Turing machine as an expression in the form of Pq_iQ where P and Q are tape expressions containing only the letters S_j, and Q cannot be empty. The position of q_i indicates that the head of the Turing machine is scanning the symbol to the right of q_i. We stipulate that a *computation of a Turing machine Z* is a finite sequence of α_i, and the last element of the sequence is called the *terminal* with respect to Z. An α_i is the terminal with respect to Z if there is no α_j such that α_j can be obtained from α_i according to Z. A Turing machine may have an infinite sequence of α_i. In which case, we say the Turing machine does not halt. Here, a computation of a Turing machine is always a finite sequence of α_i. Suppose α_T symbolizes the terminal description of a Turing machine Z. Then we define Output(α_T) as the output of Z. The value of Output(α_T) is the number of l's in α_T.

We may associate each number n with a tape expression in which there are n+1 l's. For instance, 4 = 11111, denoted as 11^3. Now let us see how to define a Turing machine of addition. Let q_1 denote the initial state, and the read/write head of the Turing machine scans the leftmost symbol of the initial instantaneous description, α_1, of the machine. The Turing machine may consist of the following finite set of quadruples: (q_1 1 B q_1), (q_1 B R q_2), (q_2 1 R q_2), (q_3 1 B q_3). In principle, there are infinitely many equivalent sets of quadruples which all perform the same function. Notice that when the Turing machine is scanning a B and in the state q_3, the function is undefined. In that situation, the machine halts. Readers may verify that when $\alpha_1 = q_1 11^x B11^y$, which represents the function (x + y), α_T becomes $B1^x Bq_3 B1^y$. Thus the output

---

machine are from Davis (1973). For simplicity, I omit the case of relative computability discussed by Davis. Alan Turing (1937) introduced his well-known Turing machine to define effective computability.

of the Turing machine is x + y as desired.

Putnam claims that a Turing machine can serve as a model for the mind. Just as the functional organization of a Turing machine can be described in terms of its logical states, the functional organization of a mind can be described in terms of its mental states.

1.1 Why Not Reductionism?

But if reductionism might be true, why bother to provide a functionalistic account of the mind? Putnam, in (1960), (1965), and (1967), argues against both behavioristic and physicalistic reductionisms. Logical behaviorism basically claims that "just as numbers are (allegedly) logical constructions out of sets, so mental events are logical constructions out of actual and possible behavior events."[4] Since mental events are logical constructions out of actual and possible behavior events, there is no need to take anything other than behavior events into consideration for building any theory of the mental. Mentalistic terms such as belief, desire, goal, intention, and any term referring to internal mental states should be eliminated from any respectable theory. This kind of behavioristic reductionism is rejected by Putnam mainly because in characterizing mental states, the reference to some other causally connected mental states is not eliminable. Mental states "are responsible for certain kinds of behavior, but only in the context of our beliefs, desires, ideological attitudes, and so forth."[5] The consequence of not taking the mental context into account is that the behavioristic input-output characterization of mental states would become either incomplete or, if complete, infinitely long. The reason is that the characterization of a mental state must necessarily involve the other mental states, and hence if a behavioristic input-output description of a mental state is finitely long, then it must contain some mental terms, otherwise, it has to be

---

[4] Putnam (1965), p. 25.

[5] Putnam (1965), p. 30.

infinitely long.

Putnam (1960) takes the logical states of a Turing machine and the mental states of a human being as analogous to the structural states of a Turing machine and the physical states of a human being. The description involving only the logical states of a Turing machine says nothing about the physical nature of these states. In the same way, the description involving only the mental states of a human being says nothing about the physical nature of these mental states. This consideration leads to the following famous thesis against physicalistic reductionism:

> THE MULTIPLE REALIZABILITY OF THE MENTAL: The same type of mental states can be realized by virtually any kind of thing; and the same type of physical states can display many different types of mental states.

According to this thesis, the relationship between types of mental states and types of physical states is a many-to-many mapping rather than one-to-one mapping. If so, then there must exist some lawful regularities at the mental level such that no law at the physical level can capture these regularities. Thus physicalistic reductionism which claims that all psychological laws can be reduced to physical laws must be false.

In Putnam's own words, the mapping between the logical states of a Turing machine and the structural states of the physical realization of the Turing machine is not one-to-one. Putnam states, "Although the machine has from the logician's point of view only the states A, B, C, etc., it has from the engineer's point of view an almost infinite number of additional 'states' (though not in the same sense of 'state') – we shall call these structural states."[6] Similarly the same thing holds in the case of the human being. The mapping between the mental states and the physical (or physiological) states of a human being is not one-to-one.

---

[6] Putnam (1960), p. 371.

1.2 Functional State Identity Theory

In Putnam (1960), Putnam does not hold the view that minds are literally Turing machines. He uses the Turing machine only as an analogy to the mind. But in Putnam (1967), he proposes a stronger thesis. Following Ned Block's terminology, let's call it Functional State Identity Theory.

> FUNCTIONAL STATE IDENTITY THEORY: The minds are probabilistic automata. That an organism is in [a] certain mental state means that the organism possesses an appropriate kind of functional organization (namely, machine table).[7]

The definition of the notion of a Probabilistic Automaton is similar to a Turing machine except that the states transitions are no longer deterministic. Turing machines, with transition probabilities 0 and 1, are treated by Putnam as a special kind of probabilistic automata. Mental states are identified with (disjunctions of) machine table states. Machine table states are functional states. Thus mental states cannot be identified with brain states which are physical rather than functional.

Two hypotheses are involved in the above version of functionalism. The first hypothesis states that a whole human being is a Turing Machine; the second one asserts that the psychological states of a human being are Turing machine states or disjunctions of Turing machine states. Eight years later, Putnam (1975) himself totally rejects these two hypotheses. Putnam has the following statement:

> Thus, if human beings have any states at all which resemble Turing machine states, those states must (1) be states the human can be in at any time, independently of learning and memory; and (2) be total instantaneous states of the human being–states which determine, together with learning and memory, what the next state will be, as well as totally specifying the present condition of the human being.[8]

---

[7] See Block (1980), pp. 171-84, and Putnam (1967), pp. 226-27.

[8] Putnam (1975). In Block (1980), p. 139.

Putnam (1975) argues that since our mental states (1) are not independent of learning and memory; and (2) are not instantaneous states, our mental states cannot be identified with Turing machine states. This argument is equally applicable to both deterministic and probabilistic Turing machines. The premises in Putnam's argument quoted above need no more elaboration because I think they are obviously true. Let's use Putnam's example. Being jealous depends on a great deal of information, learned facts, and habits. Thus the identity of that mental state requires something involving learning and memory. But the identity of a Turing machine state has nothing to do with learning and memory. How can the mental state "being jealous" be identified with a Turing machine state?

The other forceful argument against a Turing-machine-state model of the mind is Jerry Fodor's argument from the productivity of the mental states of organisms. The argument goes as follows: (1) The set of Turing machine states constituting the machine table of a probabilistic automaton is, by definition, a list; (2) the set of mental states of human beings is productive, hence there are infinitely many type-distinct mental states of any given person, therefore (3) while an automaton's set of Turing machine states can be exhaustively specified and listed, a person's set of mental states can at best be specified by finite axiomatization; therefore (4) the mental states cannot be identified with Turing machine states.[9]

Fodor mentions many other difficulties of functionalism in Putnam (1967). They will be dealt with in other sections because I think all these difficulties are equally applicable to all versions of functionalism including Fodor's. So far it is clear that the thesis that each type of mental state can be identified with a type of Turing machine state is false. As Putnam himself says, "My description qua Turing machine

---

[9] Block and Fodor (1972). In Fodor (1981), p. 94.

(machine table) and my description qua human being (via a psychological theory) are descriptions at two totally different levels of organization."[10]

Although Putnam (1975) rejects his own earlier functionalism, the following two hypotheses are preserved by Putnam and his followers such as Fodor, Gilbert Harman and other functionalists. The first one is the hypothesis that mental states could be realized in a bewildering variety of different ways; the second one is the hypothesis that an abstract description of human mental processes in terms of "mental states" can be obtained independently of the nature of the concrete realization of these mental processes. The second hypothesis constitutes the heart of functionalism since without the truth of the second hypothesis, any psychological theory would have to depend upon some physical theory and hence lose its autonomy. We will examine these two hypotheses in more details later.

1.3 Computational State Identity Theory

After 1975, Putnam's view changed frequently. But generally speaking, he is shifting to a position closer and closer to connectionism. I will explain connectionism later. The main point is that the realizing material substrate of the mind becomes more and more important in Putnam's view.

Putnam (1960) and Putnam (1967) give us two versions of functionalism, and we have found that the fundamental thesis of these two functionalisms is false. Let us look at the other more sophisticated version of functionalism proposed by Fodor. In Block and Fodor (1972), Fodor puts forward the following version of functionalism.

> COMPUTATIONAL STATE IDENTITY THEORY: Human beings are Turing machines. That a human being is in certain mental states means that the person possesses an appropriate kind of computational state.

---

[10] Putnam (1975). In Block (1980), p. 139.

A computational state of an automaton is characterizable in terms of inputs, outputs, and/or machine tables.[11] Both Putnam (1975) and Fodor (1972) maintain that mental states cannot be put into one-to-one correspondence with Turing machine states of an automaton. Fodor's computational state identity theory (CSIT) captures a different level of organization from that of the Turing-machine-states level. CSIT does avoid the two difficulties of the Turing-machine-state model mentioned by Putnam (1975). First of all, computational states are not states that a person can be in at any time because, by definition, a computational state involves proper inputs and outputs, and at least the input part is not totally controllable by the person. A person is in a computational state only if he is provided with some right inputs. Furthermore, since the definition of a computational state involves both inputs and outputs, the individuation of a computational state requires all the necessary information recorded on the memory tape and some other new information (input) to be printed on the tape. Hence an automaton's being in a computational state depends upon what the automaton has learned and what is in the automaton's memory. This avoids the first difficulty. Secondly, since a computational state involves both input and output, a computational state cannot be a totally instantaneous state in that it cannot fully determine what the next state will be and fully specifying the present condition of the person. This avoids the second difficulty. It seems that Fodor has found a plausible version of functionalism. Indeed Fodor and many other philosophers and cognitive scientists believe that computational state identity theory is the right theory of the mind. In the following section, I shall examine functionalism in general including both Putnam's and Fodor's versions.

---

[11] Block and Fodor (1972). In Fodor (1981), p. 95.

2. WHY NOT FUNCTIONALISM?

In its most general description, functionalism refers to any theory of the mind holding that mental states are functional states. According to this theory, mental states are characterized in terms of their causal relations to sensory inputs, behavioral outputs, and other mental states. It is unlike type-physicalism in that according to type-physicalism, (type) mental states are identified with (type) physical states. It is unlike behaviorism in that behaviorism does not allow the use of mentalistic terms in characterizing mental states. Some people treat functionalism as a sort of token-physicalism since although functionalism rejects the idea that there is a one-to-one correspondence between types of mental states and types of physical states, it does admit the possibility that each particular mental state must be realized by a particular physical substance. Some modern dualists, e.g., Castañeda, may gladly embrace functionalism because functionalism is ontologically neutral in the sense that functionalism does not exclude the possibility of a non-physical realization of the mental. This paper is not intended to deal with such a general position of functionalism. For example, functionalism as proposed by D. M. Armstrong and David Lewis is not discussed here because the version of functionalism that concerns us here is only the machine version of functionalism, i.e., the version which makes use of the concept of the Turing machine or other computing devices.[12]

2.1 Structural/Logical vs. Physical/Mental Distinctions

One of the major arguments used by Putnam in building his two versions of machine functionalism is the argument from the distinction of structural states and logical states on the one hand, and the distinction of physical states and mental states on the other hand. He observes that we can characterize the functional organization of a Turing machine without

---

[12] Cf. Armstrong (1968), Armstrong (1980), and Lewis (1972).

considering the nature of the structural state of the machine. The structural states are, by definition, the lower level physical states of the Turing machine. Similarly, we can characterize the functional organization of a human being without taking into account any of the organism's physical properties. This is Putnam's famous thesis of the autonomy of psychology. We may restate the above hypothesis as follows:

> HYPOTHESIS OF THE LOGICAL DISTINCTION OF THE STRUCTURAL AND THE LOGICAL: There is a clear-cut logical boundary between the structural states and the logical states of a Turing machine.

Both Putnam and Fodor need this hypothesis in order to make their machine analogy work. If the above hypothesis does not hold, then the structural properties of a Turing machine become relevant to the characterization of the functional organization of the machine. For the same reason, we then can no longer say that the physical properties of the organism are irrelevant to the functional characterization of the organism. The outcomes then are that psychology would lose its autonomy, and the thesis of the multiple realizability of the mental would lose its ground. This consequence would cheer physicalists most because the argument from the multiple realizability of the mental is the major argument against type-physicalism.

Two categories c and c′ are said logically distinct if c and c′ are not mutually reducible. c can be reduced to c′ means that all occurences of c in some description can be replaced by c′ without changing the meaning of the description. Fodor's examples is that in the case of an internal combustion engine, the camshaft plays the role of valve lifter, the "camshaft" is a structural concept while "valve filter" is a functional concept. The concept of valve lifter can be realized by many other things other than a camshaft. Thus the functional concept "valve lifter" is defined in terms of the role it plays in the case of the internal combustion engine.[13]

---

[13] Cf. Fodor (1967). In Castañeda (1967).

In both articles Putnam (1960) and Putnam (1967), Putnam explicitly states his structural/logical thesis. Even in his Putnam (1975), after criticizing his own earlier version of machine functionalism, he still says the following:

> The positive importance of machines was that it was in connection with machines, computing machines in particular, that the notion of functional organization first appeared. Machines forced us to distinguish between an abstract structure and its concrete realization. Not that that distinction came into the world for the first time with machines. But in the case of computing machines, we could not avoid rubbing our noses against the fact that what we had to count as to all intents and purposes the same structure could be realized in a bewildering variety of different ways.[14]

Putnam says that machines forced us to distinguish between an abstract structure and its concrete realization, and this distinction guarantees that the abstract structure can be realized in a great many ways.

The question is: does there really exist a logical instead of pragmatic boundary between structural and logical states? The answer is to the contrary.[15] It is well-known that in computer architecture, the boundary line between software and hardware is not fixed. Any software can be hardwired into a machine and hence become part of the hardware. Most hardware in a machine can be replaced by a program and run on a simpler machine. For example, if we consider only the functional equivalence, all jobs done by today's sophisticated mainframes can be done by a Turing machine whose architecture is far simpler than the modern computer. There is a mathematical proof for this assertion. Turing (1937) has shown that anything computable (in the intuitive sense) can be computed by his universal Turing machine. Thus we see that, from the logical point of view, the designs of most of the hardware

---

[14] Putnam (1975). In Block (1980), p. 140.

[15] Kalke (1969) held the same view as I do here with a different argument.

of modern computers are for pragmatic purposes rather than theoretical purposes if we consider only the mathematical properties of the computing machines. The tradeoff is that the program written for the Turing machine must be much more complicated and longer than the program written for a mainframe to do a certain job. The principle is, given a certain job, to decrease the complexity of the hardware, you must increase that of the software, and vice versa. It is possible to hardwire all existing programs into a computing device and hence make the machine be a very special purpose machine if doing so is desirable.

The point to be emphasized here is that the distinction between software and hardware is most pragmatic rather than logical. The two categories (of software and hardware) are not mutually irreducible. If so, we see no reasons why machines can force us to distinguish between an abstract structure and its concrete realization since what counts as the abstract structure and what counts as the physical realization is just a pragmatic decision. It has nothing to do with the existence of the structural and logical categories. If the distinction is not logical, we see no gound to assert that in characterizing the mind, the physical is totally irrelevant.

Pylyshyn (1980) also argues that the concept of Turing machine equivalence is insufficient in modeling the mind. He argues that any adequate model of the mind has to include the description of the functional architecture of the system. In Pylyshyn's words, a description of the "cognitive virtual machine" has to be specified. A cognitive virtual machine is an uninterpreted rule schema that can be exploited by an interpretation scheme to carry out intended programs."[16] Thus, it is something one level lower than programs. The cognitive virtual machine has something to do with the lower level functional organization of the physical structure. Functional architecture is said cognitively impenetrable if it is not influenceable by propositional attitudes. That is to say, functional architecture is not influenceable by higher mental functions.

---

[16] See Pylyshyn (1980), in *Behavioral and Brain Sciences* (1980) 3, pp. 111-69.

Functional architecture mostly has to do with the lower level output and input devices and their relationship to the higher level mental functions. The higher level mental functions must operate on this cognitive virtual machine, and this cognitive virtual machine in turn operates on the actual machine. Pylyshyn's point is that some properties of the physical states of the system have to be taken into consideration in constructing a model of the mind. This says the same thing as what I suggested above.

2.2 The Qualitative Properties of the Mental States

Functionalists define mental states in terms of their causal role. Hence two (type) mental states are identical if they play the same functional (causal) role. The functional role in question is defined as a relation between sensory inputs, behavioral outputs, and other mental states.

According to Putnam (1960) and Putnam (1967), mental states are identified with machine table states. Thus mental states are defined by the role they play in the causal relation among inputs, outputs, and other machine table states. Two mental states are identical if they are functionally equivalent. Thus, the principle of the individuation for mental states is their functional role in the causal relation in question. Fodor's computational state identity theory does not identify mental states with machine table states but with inputs, outputs, and a lump of machine table states. Hence, the principle of individuation for mental states in Fodor's sense is not the functional role played by machine table states but the causal role played by mental states in Fodor's sense. The content of the mental states in Fodor's sense includes the content of input and output. Thus, we may say that mental states in Putnam's sense are mental states in a narrow scope, while mental states in Fodor's sense are mental states in a wide scope. No matter how different their definitions are, they define mental states in terms of their causal roles. If two mental states are functionally equivalent, then they are said to be identical. And this causes the difficulty to be revealed below.

Conventionally, human cognition is divided into two major categories: one includes the higher functions such as thinking, doubting, believing, problem solving, reasoning, etc.; the other includes the lower functions such as feeling, sensation, emotion, perception, motor action, etc.[17] The higher functions of human cognition are usually called propositional attitudes. It seems to be less problematic to define mental states of a higher kind in terms of their functional roles. For example, a person x believes that p if the person is in a certain mental state, and he accepts some propositions as inputs, and utters the sentence 'p'. If the person x believes that q and the belief-state q plays exactly the same causal role as the belief-state p described above, then it seems to be true that the belief-state p and the belief-state q are identical. Type-identifying the mental states of higher functions poses no serious problem for functionalists. But when we consider the case of lower functions, the situations becomes very messy. Block and Fodor (1972), in criticizing the functional states identity theory, point out that it is possible for two persons to be in the same mental state but with different mental (qualitative) contents. That is to say, it is possible that functionally identical psychological states might be qualitatively distinct. For example, both you and I could be in functionally identical pain states. According to functionalism, we are in the same mental state. But it is possible that even though our pain states are functionally identical, we may, nevertheless, feel quite different things. The degree of pain I feel may be very different from the degree of pain you feel. This is the so-called "inverted qualia argument."

The situation can be even more extreme. It is possible that two persons are in two functionally identical states, e.g., pain, but only one of them really feels pain, and the other one feels

---

[17] Connectionists may disagree with this conception of human cognition. They tend to think that all cognitions are of the same nature, abiding by the same laws, functioning in the same way. More will be said in later sections on the connectionists' conception of human cognition.

nothing at all. This argument is called "the absent qualia argument." The gist of these two arguments is that functionalism fails to provide a sound account for the qualitative contents of certain mental states. Functionalists still cannot provide a good solution to this problem. I will not deal directly with the issue of qualia, but I assert that in adopting a different approach, e.g., PDP models, the qualia issue can be solved more satisfactorily.

The other related issue raised by Thomas Nagel (1974) is the issue of the subjective character of experience. In his famous article, "What is it Like to be a Bat?" Nagel argues that the subjective character of experience is fully comprehensible only from the subject's point of view. Only a bat knows what it is like to be a bat. The main purpose of Nagel's article is to show that physicalism cannot provide us with a greater understanding of human minds. I think the same argument holds for functionalism. The functionalistic account totally leaves out the subjective aspect of the mind if there is a subjective aspect of the mind at all. I do not see how functionalism can solve Nagel's problem, if functionalism itself has trouble dealing with the qualia issue.

2.3 How Can a Formal Model of the Mind Have Semantics?

Functionalists maintain that a type of psychological state is individuated in terms of its causal role. There are two basic elements in modern functionalism: one is the concept of representation and the other is the concept of computation. According to Fodor (1980), a representational theory of the mind can be stated as follows:

> THE REPRESENTATIONAL THEORY OF THE MIND: To individuate a mental state is to specify a relation and a representation so that the subject bears the one to the other.

For example, to individuate a belief state, you need to specify a relation, namely believing, and a representation, i.e., the representation of the content of the believing. Given the relation and the representation, we then can say that the subject bears the relation believing to the specified representation. Fodor further proposes a stronger doctrine:

> THE COMPUTATIONAL THEORY OF THE MIND: The representational theory of the mind is true, and mental states and mental processes are computational.

Computational processes are both symbolic and formal in the sense that they are defined over representations and they apply to representations in virtue of the syntax of the representations. Thus mental states and mental processes are said to be symbolic and formal.

That an operation is formal means that it is specified without reference to semantic properties such as truth, reference, and meaning. What is the consequence of the assumption that mental processes are formal processes defined over the internal symbolic codes? From the assumption that mental states and mental processes are formal, it follows that two mental states can be distinct in content only if they can be individuated according to formally distinct representations. Concerning the semantics of the mental states and mental processes. Fodor (1980) maintains that mental states have access only to the formal properties of the representations of the external world. This is called the formality condition:

> THE FORMALITY CONDITION: "So long as we are thinking of mental processes as purely computational, the bearing of environmental information upon such processes is exhausted by the formal character of whatever the oracles write on the tape."[18]

The above quotation is sometimes called the "completeness" assumption of the formality condition. The formality condition assures that all semantical contents of mental processes can be captured without any "semantical effort."

The formality condition is a logical consequence of the

---

[18] Fodor (1981), chapter 9, p. 231.

representational theory of the mind. Remember that one of the assumptions of representationalism is that the mind works upon the representations of the world rather than upon the world itself. The other assumption is that to individuate a mental state is to specify a mental modality (a relation) and a representation of something with which the subject is concerned, and specify that the subject bears the relation to that representation. The direct objects of the mental operations are not the things in the world but rather the representations (or ideas) formed by the subject. If so, then to construct a psychological theory is to construct a theory about the nature of mental states and mental processes, and the mechanism of how these mental states and mental processes work together. This idea is very close to what Brentano says about the existential indifference of mental phenomena.[19] Thus, to understand psychological phenomena, the concept of truth, reference, and meaning play no role in the construction of psychological theory. Fodor says that the formality condition is a property of computationalism. But, as we have seen above, it is also a logical consequence of the representational theory of the mind. A process is computational if the process can be characterized in terms of a set of computable operations and a set of representations of inputs and outputs. This is the same as saying that a process is computational if it is Turing machine characterizable. Thus a computational process, by definition, is a formal process. A process is formal if the characterization of the process involves only the physical properties, e.g. the shape of the symbols employed in the description of the process and the causal relations within the process (or among a group of the processes) involve only the physical properties of those symbols.

The formality condition, according to Fodor, is tantamount to a sort of methodological solipsism. Indeed both hypotheses

---

[19] Cf. B. Terrell, "Brentano's Philosophy of Mind," in *Contemporary Philosophy*, Vol. 4: *Philosophy of Mind.* Edited by G. Floistad (1986). Dordrecht: Martinus Nijhoff Publishers, pp. 233-34.

say something very similar. According to Putnam, the assumption can be stated as follows:

> METHODOLOGICAL SOLIPSISM: No psychological state, properly so-called, presupposes the existence of any individual other than the subject to whom that state is ascribed.[20]

If mental states are formal, then only the formal properties of the representations are relevant to the characterization of mental states. This is what the formality condition says. The methodological solipsism says that the reference of the mental representation is irrelevant to psychological state. Both assumptions claim that truth, reference, and meaning are outside the scope of psychology. Hence, according to Fodor, the semantics of mental representations is not a business for psychologists. Psychologists should be concerned only with the formal properties of mental representations.

This sort of formal approach to mental representations is sometimes called the syntactic theory of the mind (STM).[21] Stich states:

> The basic idea of the STM is that the cognitive states whose interaction is (in part) responsible for behavior can be systematically mapped to abstract syntactic objects in such a way that causal interactions among cognitive states, as well as causal links with stimuli and behavioral events, can be described in terms of the syntactic properties and relations of the abstract objects to which the cognitive states are mapped.[22]

Stich denies the explanatory function of the representational contents within the computational models of the mind. Stich is content to sketch the computational view of the mind entirely in terms of formal structures, formal rules, and functional architecture of mental representations.

Daniel C. Dennett is another famous contemporary advo-

---

[20] Putnam (1975), pp. 215-71.

[21] See S. Stich (1983), pp. 149-83.

[22] Ibid., p. 149.

cate of the syntactic model of the mind. In his widely cited paper, "Intentional Systems," he says,

> . . . intentionality is primarily a feature of linguistic entities–idioms, contexts– . . . an idiom is intentional if substitution of codesignative terms do not preserve truth or if the "objects" of the idiom are not capturable in the usual way of quantifiers.[23]

A system is an intentional system, he maintains, if the system can be explained and predicted by relying on ascriptions to the system of beliefs and desires. To explain and predict a system's behavior, Dennett argues that there are three different stances we may take. They are the design stance, the physical stance, and the intentional stance. The design stance has to do with the notion of function. We explain or predict the behavior of a system by appealing to the functions of its parts and see how these functions can serve a certain kind of intended purpose. From a physical stance, our predictions are based on the actual physical states of a particular system. The laws of nature are used in constructing prediction and explanation. Finally, from an intentional stance, "one predicts behavior in such a case by ascribing to the system the possession of certain [beliefs] and supposing it to be directed by certain [desires], and then by working out the most reasonable or appropriate action on the basis of these ascriptions and suppositions."[24] Besides the assumption of the rationality of the system, Dennett argues that we do not have to assume the existence of the beliefs and desires of the system. Intentional ascription does not assume the existence of intentionality. The rationale of adopting an intentional stance is pragmatic. If by adopting an intentional stance we may have a better prediction and explanation of the system, then the intentional stance is fully justified regardless of whether the system really possesses beliefs and desires.

For example, in predicting the behavior of a chess-playing

---

[23] Dennett (1978), chapter 1, p. 3.

[24] Dennett (1978), p. 6.

computer, an intentional stance is the best strategy to adopt even though many people may feel uncomfortable ascribing beliefs and desires to the computer because we usually do not believe that a computer can possess beliefs and desires. Dennett's position is sometimes called intentional pragmatism in contrast with intentional realism. Because intentional ascription to a system is just a matter of a pragmatic decision, Dennett (1981) further argues that for an automatic formal system with an interpretation, the semantics can take care of itself. Dennett calls the system a semantic engine. As we have seen above, an interpretation of a formal system, at the psychological level, amounts to an intentional ascription to the system. Hence, the interpretation in question does not assume the existence of the entities involved in the interpretation. This amounts to saying that the intentional existence is irrelevant to intentional prediction and explanation. And this sounds like Fodor's formality condition and Stich's syntactic model of the mind.

The evidence of the formality condition of mental phenomena, at least in the case of propositional attitudes, partially comes from the observation of the referential opacity of propositional attitudes. Fodor calls the phenomenon of opacity, "Frege's Condition."

> FREGE'S CONDITION: (THE OPACITY OF PROPOSITIONAL ATTITUDES) Sentences containing verbs of propositional attitudes are not truth functions of their complements.[25]

A sentence containing a verb of a propositional attitude is truth-functional if the truth value of the sentence is determined by the truth value of its components according to the laws of classic logic. For example, if the sentence "Mary believes that P" is true and Q has the same truth value as P does, then the sentence "Mary believes that Q" has to be true if the former sentence is truth-functional. But we see that the truth of

---

[25] Fodor (1978). In Block (1980), p. 48.

"Mary believes that John is the murderer" does not entail the truth of "Mary believes that Ben is the murderer" even though "John" and "Ben" refer to the same person. Therefore, in charaterizing Mary's belief state, the reference of the terms in the representation "John is the murderer" is irrelevant. That is one of the reasons why Fodor, among others, claims that semantics is not the business of psychology.

The syntactic model of the mind advocated by Fodor, Dennett, Stich, and others presupposes what Tyler Burge (1986) calls individualism of intentional mental states, an individualism about the individuation of mental kinds. According to Burge (1986),

> According to individualism about the mind, the mental natures of all a person's or animal's mental states (and events) are such that there is no necessary or deep individuative relation between the individual's being in states of those kinds and nature of the individual's physical or social environments.[26]

Burge (1986) argues that this assumption is implausible. Furthermore, not all psychology as it is currently practiced is or should be individualistic.[27] One strong argument supporting individualism is Putnam's Twin Earth analogy.[28] Putnam argues that it is possible for two persons to be in exactly the same psychological state, even though the extensions of the term used by both persons are different. Extensions are not determined by psychological state. Hence, Putnam claims that meanings are not in the head! By assuming the individualism of mental states, Putnam can claim that semantics are irrelevant. But with a non-individualistic view of mental states, it becomes questionable if it is really true that meanings are not in the head. The connectionist approach to be defended in

---

[26] Tyler Burge (1986), pp. 3-4.

[27] I will not repeat Burge's lengthy argument here. For more details, please see Burge (1979), Burge (1986), and "Other Bodies" in Andrew Woodfield (1982), pp. 97-120.

[28] Putnam (1975), pp. 215-71.

this paper assumes a non-individualistic view of the mind in that the connectionist models take both the individual's physical and social environments seriously into account.

Another important issue concerning the semantics of mental representations is the failure of Fodor and others to distinguish procedural semantics from the conventional Tarskian type semantics. According to the machine models of the mind, mental processes are treated as analogous to computer programs. The semantic issue of computer programs has two aspects: one is the issue of how a computing machine interprets the codes and does whatever the codes instruct; the other one is the issue of interpreting what the machine is doing at a higher level of description when the machine is executing the instructions. The former kind of semantics is called procedural semantics. The procedural semantics of a program, in the case of a computer, is accomplished by a compiler or assembler. The compiler or assembler translates the programs written in higher level languages into, eventually, the machine-executable codes, i.e., machine language. This is the process of a machine's interpreting the "meaning" of a program. A program works only if it is machine executable, i.e., only if it can be translated into proper machine codes. The second kind of semantics is a sort of conventional semantics whose purpose is to treat a program as a model of the behavior of the system. In understanding what a program is doing, according to the second sort of semantics, we need to assign references to all the terms employed in the program, and hence provide a semantic evaluation of the program. Without the consideration of reference, truth, and meaning, a program is simply a formal chunk of symbols and can mean anything. For example, it is possible in artificial intelligence programs to find that exactly the same set of codes can be interpreted as a program for chess-playing and as a program for a star wars video game. How the program is interpreted depends upon the semantic content of the codes. Hence the individuation of the program seems to involve the content of the codes which goes counter to the formality assumption of the syntactic models of the mind. Fodor and others try to provide a causal mechanism of mental processes by adopting machine models and procedural semantics to guarantee the physical realizability of the mental, but they are wrong in confusing the procedural semantics with conventional semantics. It is certainly true that in constructing procedural semantics for mental representations, the involvement of the external environment is irrelevant. But to understand what is going on in a mental phenomenon, conventional semantics is not avoidable.

2.4 The Fundamental Assumptions of Machine Models of the Mind

In this section, certain basic assumptions of machine models will be examined. The machine models of the mind include Turing machine models and von Neumann type machine models. The architecture of a von Neumann machine consists of a set of registers which serve as the working memory of the machine, a central processing unit (CPU) containing a control unit and an arithmetic unit, a secondary memory which serves as permanent storage for memory, and some input and output devices. A Turing machine, theoretically speaking, is a specialized von Neumann machine with only two registers (read register and write register), without the distinction of the working memory and secondary memory. The number of registers a von Neumann machine can have is theoretically unlimited even though in a realistic von Neumann type machine, there are usually only a limited number of registers. When there are an unlimited number of registers, the von Neumann machine is called the unlimited register machine (URM).[29]

There are four major assumptions in the von Neumann machine model: (1) the psychological processes are formal and symbolic; (2) psychological behaviors are rule-governed rather than rule-described; (3) the psychological processes are computed in a sequential (serial) manner, one instruction at a time

---

[29] Cf. Cutland (1980), pp. 9-24.

rather than a massively parallel manner; and (4) the memory storage is local and rigid rather than distributed. All of these assumptions will be examined and rejected in the following sections.

2.4.1 The Symbolic Paradigm

Two components of a symbol system are (1) a set of primitive symbols; (2) a set of rules for forming derivative symbols and a set of rules for determining the interconnections of these symbols. Generally speaking, the von Neumann model of the mind claims that psychological systems are symbol systems. Fodor (1975) argues that the best psychological theories are computational theories; but there is no computation without representation, therefore there is no representation without a symbol system (or language of thought).[30] Paul Smolensky (1988) describes the symbolic paradigm as follows:

> THE SYMBOLIC PARADIGM: Cognitive descriptions are built of entities that are symbols both in the semantic sense of referring to external objects and in the syntactic sense of being operated upon by "symbol manipulation."[31]

The major argument Fodor uses to support the assumption of the symbolic nature of representations is the argument of the systematicity of representations.[32] The nature of systematicity says that if the representational capacity of a language can represent a proposition, then the other propositions which are semantically related to the former are also representable by the language. For example, if "Mary loves John" is representable, then "John loves Mary" is representable too. The systematicity of a representational system implies that the representations of a system form a generated set. A set of representations is a generated set if the set can be defined in an induc-

---

[30] Fodor (1975), pp. 31-32.

[31] Smolensky (1988).

[32] Fodor (1987), pp. 135-54.

tive (or recursive) way. Only a symbol system has internal syntactic and semantic structures which are necessary for a generated system. The reason is that the constituency relation between parts and whole is available only in a system with internal syntactic and semantic structures.

The connectionist models of the mind denies that psychological processes are symbol manipulation processes. Connectionists propose a kind of subsymbolic paradigm, as Smolensky calls it. The subsymbolic paradigm asserts that cognitive descriptions are built with constituents of the symbols used in the symbolic paradigm. Fodor maintains that the constituents of a symbol are still symbols. But connectionists want to go down one more level and treat the subsymbolic entities as the objects of mental operations.

There are a great variety of connectionist models. However, generally speaking, all connectionist models must include the following two basic features: distributed representation and numerical (or non-symbolic) computation. I have elsewhere argued that a non-distributed or localist representation scheme will make a connectionist model become just an implementation of the symbolic architecture.[33] On the other hand, if mental processes operate in virtue of the syntactic structures of symbolic representations, then there is no need to go down one level and specify the processing mechanisms of the mind in terms of the numerical properties of the subsymbolic units. Therefore, if the connectionist approach is intended to be a genuine alternative approach to the symbolic approach, the above two features are the minimal requirement.

2.4.2 Rule-Governed vs. Rule-Described Behavior

Gilbert Ryle, criticizing a so-called intellectualist legend, says, "Champions of this legend are apt to try to reassimilate knowing how to knowing that by arguing that intelligent performance involves the observance of rules, or the application

---

[33] See Houng (1991), chapter 1.

of criteria."[34] The von Neumann model of the mind assumes precisely a form of intellectualist view of cognitive behavior. Modern computational models of the mind hold that all of our intelligent behaviors are rule-governed and for some class of rule-governed behaviors, we even have no conscious knowledge of the rules. The syntactic rules of our natural language are frequently cited as an example. In language production we all follow syntactic rules, perhaps a sort of universal syntax, but few of us possess a conscious knowledge of the syntax of our language.

Ryle (1949) argues against this intellectualist assumption to maintain the distinction of know-that and know-how. The other problem that worries Ryle is that an intellectualist assumption will leads to a vicious infinite regress. The infinite regress results when the application of a set of represented rules requires another set of rules to regulate the application of the first rules. Then you need the third set of rules to regulate the application of the second set of rules, and so on and so forth. The infinite regress has been shown by the modern computer to be avoidable. The point is that, in the case of the computer, eventually the rules are translated into primitive instructions and directly executed by the hardwired circuitry, hence, terminating the regress.

From the connectionists' point of view, Ryle is justified in distinguishing know-that from know-how. According to machine models of the mind, all mental operations are rule-governed. A computer program consists of a set of instructions. When running a program, a computer executes each instruction one at a time. Machine functionalists believe that mental operations are just like computer programs. But consider skill learning. Hubert Dreyfus and Stuart Dreyfus (1986) divide skill learning, chess-playing in particular, into five stages: novice, advanced beginner, competent, proficient, and expert. In learning chess-playing, most beginners attempt to remember all the rules and their priorities. When they become

---

[34] Gilbert Ryle (1949), p. 29.

advanced beginners, they begin to recognize such situational aspects of positions as a weakened king's side or a strong pawn structure despite the lack of precise and universally valid definitional rules. At the competent stage, they learn to adopt a hierarchical view of decision-making in order to cope with the information explosion. When they enter the stage of proficiency, they can recognize a large repertoire of types of positions without too much conscious effort, but they still must deliberate about how best to do so. At the expert stage they do not make conscious deliberative decisions. Except during moments of breakdown, they understand, act, and learn from results without any conscious awareness of the process.[35] The same observation holds in the case of talking, riding a bike, and driving a car. An expert car driver would not consciously follow a set of rules to drive a car. Machine functionalists may reply that the driver follows the rules unconsciously. But then machine functionalists owe us an explanation of why the rule-following becomes unconscious when a person becomes an expert in some field. Functionalists may answer that the rule-following has to become automated and hence unconscious. It is true that the run time of an automated subroutine (compiled subroutine) is shorter. But it also means that the revision and change of the compiled subroutine is much more difficult than a subroutine written in a higher language (the source codes). This computer analogy shows that if an experts, says a chess player, has to follow the rules unconsciously, then any improvement in his expertise would become very difficult. That means that it is more difficult for an expert to learn new chess skills than it would be for a beginner. This seems false.

That a phenomenon is rule-describable does not imply that the phenomenon is following the rule. For example, there are rules (physical laws) which describe the phenomenon of our Earth's circling around the Sun. But it does not follow that the Earth either consciously or unconsciously follows the rules.

---

[35] Dreyfus & Dreyfus (1986b). In N. E. Sharkey (1986), chapter 12, pp. 315-35. Also see Dreyfus & Dreyfus (1986a).

Kripke (1982) shows that given any sequence of natural numbers, you may formulate many different rules according to which the sequence is constructed in many different ways. Even if mental operations are a rule-following business, we can still question the psychological reality of the rule-following.

The hypothesis of rule-governed mental behavior presupposes the symbol system hypothesis. All rules are formulated in a kind of language. Fodor claims the existence of the language of thought, and that the rules are written in this language of thought. Any language, properly speaking, has to have syntactic and semantic structures. Syntactic structures require the use of symbols. But if we adopt a connectionist approach, denying that mental operations operate upon a set of symbolic entities, then the hypothesis of the rule-governed nature becomes untenable.

2.4.3 The Hypothesis of Sequentiality

The von Neumann type digital computer reads one thing at a time from memory and executes one instruction after another. The computational processes are performed in a sequential or serial manner. The idea of parallel processing has been implemented on today's digital computers, but there still exists an essential difference between limited parallel processing and authentically massive parallel processing which is employed by our neural system. Each neuron is regarded as a processing unit. The number of neurons in our neural system is far more than the number of CPU's in any parallel digital computer. Even if we could build a massively parallel digital computer, the other hypotheses of the von Neumann model discussed in this section prevent us from building a genuine neuronal computing machine.

This essential difficulty of the sequential machine is its speed. As indicated by Dana H. Ballard (1986),

> One of the deepest mysteries of the function of the cortex is that neural processing times are only about one hundred times as fast as the fastest response times for complex behavior. At the very least, this would seem to indicate that the cortex does

massive amounts of parallel computations.[36]

For serial computers, the basic operations are measured in nanoseconds, but for neurons, the basic operations are measured in the milliseconds. Brain is x • 10 times slower than serial computers.[37] Ballard says that the neural processing times are only about 100 times as fast as the fastest response times for complex behavior. But we are able to do very sophisticated processing in a few hundred milliseconds. These tasks must be done in no more than 100 or so serial steps. This is the so-called 100-step program constraint.[38] Therefore, a massively parallel processing mechanism is required for coping with the problem of speed.

Fodor (1983) proposes the concept of the modularity of mind. The major idea is that many fundamentally different kinds of psychological mechanisms must be postulated in order to explain the fact of mental life. There are, Fodor contends, many different psychological faculties (the modularity thesis, such as the linguistic module, visual module, motor module, etc. Each module functions independently without communication among them. All modules are connected with central systems. The central systems use the same language, for example, the language of thought. All modules may use different representational systems, but all of their outputs are compiled into the representations used by the central systems. Fodor's modularity thesis suggests a kind of limited parallel processing. It can increase the speed of the processing by allowing each module to work independently and simultaneously. But the speed constraint is so tight that the modularity thesis provides little help. A cooperative activity among a huge number of processing units seems to be mandatory.

Notice that Minsky (1985) also proposes a sort of cooperative computation. His main thesis is to reject the *single self* assumption in psychology, and put forward a *multiple self* thesis he calls *the society of mind.* According to Minsky (1985), there are large collection of simple minds, and the high level

---

[36] Ballard (1986) in *The Behavioral and Brain Sciences* (1986) 9, pp. 67-120.

[37] See Rumelhart, McClelland, and the PDP Research Group (1986), p. 130.

[38] Feldman (1982).

intelligent behavior of a person is realized by the cooperative activities of the society of mind. Besides Minsky, researchers in traditional artificial intelligence now also recognize the advantage of the massively parallel computation. But Minsky and many other AI people who emphasize parallel computation are nevertheless still different from connectionists in that they do not give up the assumptions of symbolic representation and syntax-sensitive computation.

2.4.4 The Local vs. Distributed Conceptions of Memory

There are some psychological phenomena for which any model of the mind has to provide a plausible and coherent explanation. The first phenomenon is so-called graceful degradation with damage and information overload.[39] Graceful degradation refers to the fact that our neural system's performance gradually deteriorates as more and more neural units are damaged, but there is no single critical point where performance breaks down. Alzheimer's disease provides the best example of graceful degradation.

The von Neumann model of the mind is not able to provide a plausible explanation for the phenomenon of graceful degradation. In a computer program, the damage or change of a single bit can catastrophically deteriorate the performance of the program. The degree of toleration of the damage to the system is very low in the case of the von Neumann model of the mind.

The second psychological phenomenon is so-called content addressability.[40] Human memory is content addressable in the sense that we can retrieve information from our memory based on nearly any attribute of the representation we are retrieving. For example, suppose you want to access a piece of information about your grandmother. You may start with any memory

---

[39] The materials used in the following discussion primarily come from Rumelhart & McClelland, "PDP Models and General Issues in Cognitive Science," in Rumelhart, McClelland, and PDP Research Group (1986), pp. 110-46.

[40] See McClelland, Rumelhart, and Hinton, "The Appeal of Parallel Distributed Processing," ibid., chapter 1, pp. 29-30.

about your grandmother available to you at this point, and finally retrieve the piece of information you want. In this case, you use a piece of information as a clue to access the other piece of information. In other words, you use the content of certain information to address the content of the other piece of information. This kind of property of mental phenomena is called content addressability.

In the case of the von Neumann computer, programs are written so that if you want to access a piece of information, you have to know the address of the piece of information. Otherwise, the information is simply not accessible. You are not able to address a piece of information by the use of another piece of information. The only way you can obtain the desired information is to know the address of the piece of information which you want to access.

The third psychological phenomenon is the so-called perceptual completion of familiar patterns and the completion of novel patterns.[41] It is a well-known fact that we can get by with less time or with lower-quality information in perceiving a familiar object. If the object to be perceived is familiar to us, then the information needed to perceptually identify the object can be relatively poorer than that needed in perceiving an unfamiliar object. Even in the case of a novel object, "we also show facilitation in the perception of letters in unfamiliar letter strings which are word-like but not themselves actually familiar."[42] The ability to fill in missing portions with imperfect perceptual inputs has to be explained in any model of the mind.

Due to the rule-governed nature of the von Neumann model, the perceptual completion of both familiar and novel patterns is achieved by pattern matching. But pattern matching requires a detailed set of rules to take care of all the relevant aspects in a perceptual completion task. The missing portions of perceptual input cannot be too large, otherwise the program may match the input to a too large set of possible

---

[41] Ibid., pp. 20-25.

[42] Ibid., pp. 23-24.

candidates. This makes pattern matching very inefficient and useless in perceptual identification. It has been shown that an ordinary human perceptual completion task is just too difficult for a von Neumann machine to model.

The incapability of the von Neumann model in dealing with the issues discussed above lies in the particular conception of memory they adopt. The scheme of representations used in the von Neumann model is a sort of local representation scheme. A local representation scheme uses one computing element for each entity. Information is stored in some place and to retrieve it, you have to go to that specific location to find it. In general, memory is stored by place. For example, in a digital computer system, a piece of information is stored at a certain address. Each address stores one item of information. On the other hand, a distributed system stores a piece of information by a pattern of activity distributed over many computing elements. The same set of computing elements may be used at the same time to represent different entities, and the same entity is represented by a set of different computing elements.[43]

The distributed conception of memory maintains that information is not stored at a particular place, rather it is stored by the pattern of the connectivity of a set of computing units. A distributed system allows you to create a new concept without allocating new hardware in that the same set of computing elements can be used to represent many different concepts. All you need to do is to modify the interaction between units to create a new pattern of activity for the new concept. Of course this new pattern of activity must not interfere or change the old, possibly many, patterns of activities, and this poses no problem for distributed models. In a conventional local representation scheme, to form a new concept you need to allocate a spare memory location for the concept.

In a distributed system, semantically related information

---

[43] See Hinton, McClelland, and Rumelhart, "Distributed Representations," in ibid., chapter 3, pp. 77-109.

tends to evoke one another. Thus, it is easy to model content addressability in a PDP model. Because the memory in a PDP model is stored by the pattern of the activity of some set of computing units, damage to a number of units usually does not prevent the pattern of the activity from being evoked. This makes graceful degradation of memory possible in a PDP model. Finally, since semantically related information tends to evoke one another, even if only a very low quality of input is provided, the initial input would evoke other semantically related items and form a set of constraints to regulate the evoking of the remaining part of the pattern. This makes the perceptual completion task easy to accomplish in a connectionist model.

3. CONCLUSION

I have examined three different versions of machine functionalism, two proposed by Putnam and one by Fodor and many others. All forms of functionalism discussed here are rejected. I have argued that the structural vs. logical distinction is primarily pragmatic rather than theoretical, and hence, the distinction between the physical and the mental also lacks theoretical ground. I have also argued that functionalism cannot offer a satisfactory explanation of qualitative mental properties and cannot explain why a formal model of the mind can have semantics. Finally, I have argued that four basic assumptions, i.e., the symbol system, the rule-governedness, sequentiality, and the localist conception of memory, of machine functionalism are false and indicated that connectionist models as alternatives seem to be more promising in attacking the above difficulties.

REFERENCES

Armstrong, D. M. 1968. *A Materialist Theory of Mind.* London: Routledge & Kegan Paul.

———. 1980. *The Nature of Mind and Other Essays.* Ithaca, N.Y.: Cornell Univ. Press.

Ballard, D. H. 1986. "Cortical Connections and Parallel Processing: Structure and Function." *Behavioral and Brain Sciences* 9 (1986): 67-120.

Block, Ned, ed. 1980. *Readings in Philosophy of Psychology.* Vols. 1 & 2, Cambridge, Mass.: Harvard Univ. Press.

Block, Ned and Jerry Fodor. 1972. "What Psychological States Are Not." *Philosophical Review* 81 (1972): 159-81. Also in Block (1980): 237-50.

Burge, Tyler. 1979. "Individualism and the Mental." In French, et al. eds. (1979).

———. 1986. "Individualism and Psychology." *The Philosophical Review* (1986): 3-45.

Castañeda, H. N., ed. 1967. *Intentionality, Minds and Perception.* Detroit: Wayne State Univ. Press.

Cutland, N. J. 1980. *Computability*. Cambridge Univ. Press.

Davis, M. 1973. *Computability and Unsolvability.* New York: Dover.

Dennett, Daniel. 1978. *Brainstorms.* Montgomery, Vt.: Bradford Books.

Dreyfus, H. L. and S. E. Dreyfus. 1986a. *Mind over Machine.* New York: The Free Press.

———. 1986b. "Why Skills Cannot Be Represented by Rules." In Sharkey (1986), pp. 315-35.

Feldman, J. A. 1982. "Connectionist Models and Their Properties." *Cognitive Science* (1982): 205-54.

Fodor, Jerry. 1967. "Psychological Predicates." In Castañeda (1967).

———. 1975. *The Language of Thought.* New York: Thomas Y. Crowell.

———. 1978. "Propositional Attitudes." *Monist* 61 (1978): 501-21.

Fodor, Jerry. 1981. *Representations: Philosophical Essays on the Foundations of Cognitive Science.* Cambridge, Mass.: The MIT Press.

———. 1987. *Psychosemantics.* Cambridge, Mass.: The MIT Press.

French, P., T. Uehling, and H. Wettstein, eds. 1979. *Midwest Studies in Philosophy*, Vol. 4: *Studies in Metaphysics.* Minneapolis: Univ. of Minnesota Press.

Hook, Sydney, ed. 1960. *Dimensions of Mind: A Symposium.* New York: New York Univ. Press.

Houng, Y. 1991. *Classicism, Connectionism, and the Concept of Level.* Ph.D. dissertation, Indiana University.

Kalke, W. 1969. "What Is Wrong with Fodor and Putnam's Functionalism?" *Nous* 3 (1969): 83-93.

Lewis, D. 1972. "Psychological and Theoretical Identifications." *Australasian Journal of Philosophy* 50 (1972): 249-58.

Loemker, L. E. ed. 1976. *Leibniz: Philosophical Papers and Letters.* Syntheses Historical Library.

Minsky, M. 1985. *The Society of Mind.* New York: Simon & Schuster.

Nagel, T. 1974. "What Is It Like to Be a Bat?" *Philosophical Review* 83 (1974): 435-50. Reprinted in Block (1980).

Putnam, H. 1960. "Minds and Machines." In Hook (1960).

———. 1965. "Brain and Behaviors." In Block (1980), Vol. 1, pp. 24-36. Originally in R. J. Butler (ed.), *Analytical Philosophy*, Vol. 2. Oxford: Blackwell, 1965.

———. 1967. "The Mental Life of Some Machines." In Castañeda, ed. (1967).

———. 1975. "Philosophy and Our Mental Life." In Block (1980), p. 139. Originally in Putnam, *Mind, Language, and Reality: Philosophical Papers*, Vol. 2. London: Cambridge Univ. Press, 1975.

Pylyshyn, Zenon. 1980. "Computation and Cognition: Issues in the Foundations of Cognitive Science." *Behavioral and Brain Sciences* (1980): 111-69.

Rumelhart, D. E., J. L. McClelland, and the PDP Research

Group. 1986. *Parallel Distributed Processing: Explorations in the Microstructures of Cognition.* Vols. 1 & 2. Cambridge, Mass.: The MIT Press.

Ryle, G. 1949. *The Concept of Mind.* New York: Barnes and Noble.

Searle, John. 1984. *Minds, Brains and Science.* Cambridge, Mass.: Harvard Univ. Press.

Sharkey, N. E., ed. 1986. *Advances in Cognitive Science 1.* New York: Ellis Horwood Limited.

Smólensky, Paul. 1988. "On the Proper Treatment of Connectionism." *Behavioral and Brain Sciences* 11 (1988): 1-74.

Stich, S. 1983. *From Folk Psychology to Cognitive Science.* Cambridge, Mass.: The MIT Press.

Turing, A. 1937. "On Computable Numbers, with an Application to the Entscheidungsproblem." In *Proceedings of the London Mathematical Society* 42 (1936-37): 230-65.

Woodfield, A. 1982. *Thought and Object.* Oxford: Oxford Univ. Press.

心靈哲學中的功能論及其理論上之困難

洪　裕　宏

摘　要

功能論（Functionalism）主張我們可以用心靈狀態的因果角色來界說心靈狀態。換言之，我們可以用某一心靈狀態與感覺刺激、行為、和其他心靈狀態之間的因果關係來界說該一心靈狀態。有一派功能論主張心靈狀態的因果角色可用涂林機（Turing Machine）的概念來界說。本文認為這派功能論的許多假設和基本假說根本不能成立。本文討論的四個基本假設是：(1)認知元體在性質上是符號的；(2)認知行為是依據一套規則來運作的；(3)心理過程的進行方式是直續式的；(4)記憶系統的儲存方式是區位性的，而非分散性的。本文反對認知系統有結構的與邏輯的之區分，也反對把認知系統當做一種形式系統。因為上述的假設和假說都是錯的，所以功能論無法解決認知科學中的許多重要問題。本文認為聯結論（Connectionism）可能是一個比較可取的理論架構。聯結論強調任何合理的關於人類心靈的理論都需考慮人的大腦的構成方式。電腦與人腦在結構上差異太大，所以以電腦類比為基石的功能論自然不是一個好的理論架構。