Theory of the Mind Under Question

Sam Leventhal

Theories of the mind such as identity theory and functionalism pose the question as to whether or not the mind is fundamentally algorithmic. By introducing Godel's incompleteness theorem (GIT) we will be able to bring into question the mind's ability for accurate logical deduction and theorizing about the mind as a whole. Godel's results explain that a consistent (meaning affirmative and negative statements can not be proven simultaneously) logical system can not prove all statements of its own creation nor prove itself to be a consistent system using it's own logic. First questioning the validity of identity theory (I.T.) followed by showing a functionalist theory must accommodate for self reflection will allow us to use Godel's theorem to set limitations on the theory of the mind as a scientific practice depending on whether or not functionalism should be considered true or false.

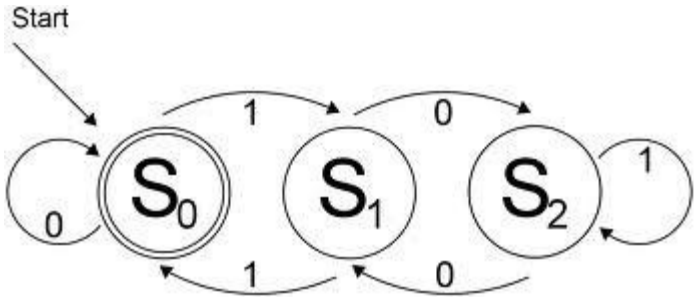
Identity theory uses the brain as a predictable biological mechanisation for all mental states, interpreting the mind's “I” as a culmination of mental states which are equivalent to a specific set of brain states. Due to mental states being dependent on the literal mechanics of the brain it can be shown that identity theory (I.T.) is either incorrect or is an unilluminating theory of the mind. If mental states are in fact directly tied to the mechanics of the physical brain we must ask how differing brains can produce identical mental states. I.T. must then resolve the question as to if the human's mental state of pain is equal to the pain experienced by an octopus despite the fact an octopus's brain is distributed throughout its tentacles. As Putnam explains:

> "Any organism (not just a mammal) is in pain if and only if (a) it possesses a brain of suitable physical-chemical structure; and (b) its brain is in that physical-

> chemical state. This means that the physical-chemical state in question must be a possible physical state of a mammalian brain, a reptilian brain" (Putnam ‘The Nature of Mental States’, 77).

The dependence between mental states and specific brain states is then forcing I.T. to become type specific, where theorizing about the mind requires one to treat each type of brain (octopus, alien, or weevil) as possessing a specific type of mental state. If however there exists a physiological state equivalent between say, both humans and octopi, but separate in physiological structure I.T. would be refuted. However assuming I.T. to be true and ignoring the intuitive notion that fear is fear regardless of the species then I.T. will still be inadequate since a type specific theory of the mind would require a case by case analysis. I.T. is then reduced to an extremely uninformative and non-generalizable theory of the mind.

Functionalism resolves the emergence of type specific I.T. by using the algorithmic Turing machine model within computer science to explain shared mental states between multiple types of brains. Functionalism accomplishes this by exploiting the fact that the expressability of Turing machines is non-committal to the exact physical nature of the mind, allowing for mental states to be accessible by multiple types of physiology's. To quickly illustrate the abstraction used we consider deterministic finite autonomous (DFA) which can be considered a simplified Turing machine. The purpose of DFAs are to read input and assert whether or not the given input satisfies some condition. Graphically a DFA is a set of nodes connected by paths. The paths between each node act as "doors" between states and the "keys" are the input. A DFA which generates all numbers divisible by three would read numbers expressed as binary. Such a DFA would report true when given: 11 (3), 110 (6), 1001 (9), 1100 (12), 1111 (15)... Exploiting the emerging pattern results in the following DFA:



If reading through a given number in binary and moving through the graph accordingly leaves you back on S0, then the input number is divisible by three. For the functionalist mental states *are* Turing machines. As such mental states may be expressed as finite and consistent logical systems. Abstracting the functionality of the brain to a propositional system leads mental states and the brain to be similar to software and hardware, i.e. allowing the same software to be recognized on various types of hardware.

Functionalism having introduced the possibility of the mind as expressible through a logical semantics allows us to better understand the limitations of the mind in its ability to formulate valid logical arguments. Two models of functionalism exist which differ in their view as to what the mind is able to use in order to form logical assertions. In terms of Turing machines these two models differ in what they allow as input. More specifically the realizer model explains the “Turing machines” of the mind can only use basic types such as enviromental experiences. The mental state of pain would then have used C-fiber activation as it's mechanism of action. Next is role functionalism which is equivalent to Turing machines able to use other Turing machines as input. Role functionalism then allows mental states to arise from relations of lower level functionalities. The role functionalist view of pain would then consider the pain's mental state to be the functional representation able to process all other functional representations of pain. The major difference between both models is that role functionalism

incorporates realizer functionalism in that role functionalism groups mental states and their equivalent state machines. What is important to note is that role functionalism introduces the notion that there are mental states of the mind able to make logical assertions about other mental states.

Realizer functionalism's dependance on physical states causes it to permit non-cognoscente minds as well as reduce to a type specific theory. By considering mental states to be the exact mechanics of material designs such as pain being the result of C-fiber activation, we loose generality a similar manner to type I.T. Ned Block demonstrates this via a thought experiment called the "Homunculi-headed Robot" in which the a human mind is operated by a group of micro-men. By managing input identically to that of neurons, micro-men "as a whole manage to simulate you because the functional organization they have been trained to realize is yours" (Troubles with Functionalism, Block). The collective action of the micro-men results in a functional duplicate of the unique individual; a notion theories of the mind hope to preserve. Although providing the identical functional organization needed for a given mental state, the micro-men are unconscious as a whole and lack the qualitative states which you, the person they are mimicking, experience. The admittance of non-cognoscente forms, as argued by Block, results from functionalism's criteria that finite state machines are unable to analyze other logical systems, such as the micro-men's inability to reflect on themselves as a whole. In a similar vein Hilary Putnam argues functionalism to be an ambiguous form of identity theory where mental types are instead equivalent to functional types. To demonstrate Putnam uses the hypothetical 'Twin Earth' argument: If a twin earth existed where the only "difference between the two planets is that there is no water on Twin Earth" but an alternate equivalent then all doppelgangers would experience and treat twin earth's water identically, leaving no identifiable difference between

earth and twin earth. Putnam argues that since every person and his twin have the same psychological and physiological states when considering water, despite their respective waters only being equal in subjective properties and not actual composition, demonstrates functional states of the brain as insufficient in fully explaining the concepts which they try to express. Thomas Nagel takes a more extreme approach through a theoretical 'zombie Parallel-universe' where each person has a zombie counterpart able to act identically to the earth individual, despite being zombies and having no consciousness. A common attack against realizer functionalism; as Block, Putnam, and Nagel have shown, is to construct hypothetical examples where the behavior of a given mental state is exhibited in a manner making the mechanics and sentience of the brain irrelevant.

To investigate the implications of role functionalism we must detour to explain Gödel's incompleteness theorem which places scope on the abilities of any computable system. Gödel's first proof of the incompleteness theorem shows that any consistent axiomatic language with a finite number of states will be inherently limited in expressability. A language can form concepts which may be true but can never be proven using its own semantics. What is more, Gödel continues to show that such a language is also unable to determine it's own consistency. As an example consider an Omnipotent Truth Machine (OTM). The OTM is able to say whether any statement is true of false. If Gödel first asks the OTM for it's program (P) and circuit design then Gödel will be given two finite articles, no matter how complex. Gödel then builds the following statement (G) to test against the OTM: "Any machine designed to accommodate P will never say this statement is true". If Gödel then asks OTM "Is G true?" a logical tongue twister follows. Within computer science, and therefore for a Turing machine type logic, an example of undecidability is the Halting Problem which explains the statement "given some expression of a

computer program, decide whether or not when given some input IO if the program will finish or run forever" is undecidable, i.e. can not be proved.

Objections made by opponents share in common that the mind must be able to self reflect. This however does not negate the prospect of the mind as a computational system but rather adds a condition as to what type of logical systems must express the mind. Role functionalism would be satisfactory due to its generalized scope and allowance of computational engines capable of interpreting other logical systems.

The emergence of role functionalism, regardless of its validity, places philosophy of the mind on a concerning frontier. If functionalism is true then Gödel has shown us the mind's equivalent functional representation is incomplete and unable to prove itself as capable of obtaining valid, non-contradictory ideas. As a result the science investigating the theory of the mind reaches a paradoxical dead end in a way similar to how "we cannot see our faces with our own eyes." (Hofstadter, Gödel, Escher, Bach). If functionalism is false then the mind is not a consistent and logical system. As a result the mind's ability of understanding human consciousness would be an incomplete pseudo-computational language liable to Orwellian "doublethink" leaving the mind able to accept contradictory ideas.

Works Cited

(1) Block, Ned (1978). "Troubles with Functionalism" *Minnesota Studies in The Philosophy of Science* **9**: 261–325.

(2) Hofstadter, Douglas R. (1999) [1979], Gödel, Escher, Bach: An Eternal Golden Braid, Basic Books.Language and Reality. Cambridge University Press.

(3)Nagel, "What is it Like to Be a Bat?" (1974), p. 436.

(4)Putnam, H. (1975/1985) "The Meaning of Meaning." In Philosophical Papers, Vol. 2: Mind,