The following paper was submitted for a Philosophy class on Intermediate Logic. It concerns Kurt Gödel's philosophy of mind as it relates to his Second Incompleteness Theorem.

When considering the philosophical implications of his Incompleteness theorems, Kurt Gödel argues for the following disjunctive statement: "either mathematics is incompletable in this sense, that its evident axioms can never be comprised in a finite rule, that is to say, the human mind (even within the realm of pure mathematics) infinitely surpasses the powers of any finite machine or else there exist absolutely unsolvable Diophantine problems of the type specified." In what follows, I begin by sketching Gödel's argument for this elaborate statement, seeking to explain what exactly is meant by it. I then proceed to show how Gödel uses this statement to argue for the fact that if its first disjunct is true, then the human mind cannot be reduced to the workings of the human brain. Finally, I take a stance against Gödel's argument, for it explicitly assumes that the brain is reducible to a Turing Machine, a hypothesis of which I am quite skeptical, given the inherent complexity of the human brain.

It is important to first clarify what Gödel has in mind when he speaks of the inexhaustibility of mathematics. In this context, Gödel is interested in mathematics proper, which concerns itself solely with absolute truths. Hence, unlike Euclid's fifth postulate, for example, the axioms of mathematics proper are not arbitrary assumptions used to derive contingent theorems. Rather, these axioms are absolutely true and, as such, must be 'evidently true'. Gödel further distinguishes between objective mathematics and subjective mathematics, the former of which contains all true mathematical propositions, while the latter contains all demonstrable mathematical propositions. Now a formal system T is characterized by some finite set of symbols, called the language of T, along with a set of axioms and a set of logical inference rules, from which its theorems can be derived. A formal system is consistent if no contradiction can be derived from its axioms. A formal system is well-defined if its axioms, if they are finite, can be formally written down and, if they are infinite, can be produced by some finite procedure which writes them down one after the other. Hence, a formal system is inexhaustible is no such procedure exists.

It is also important to establish two key equivalences used by Gödel throughout his argument. Recall that a Turing Machine is an idealized mechanism consisting of a head moving through an infinite discrete tape over a discrete set of time instances. At any given time, the machine follows a particular instruction, and can either write down a one, write down a zero, move left, or move right. Turing Machines and well-defined formal systems can be shown to be equivalent concepts. Indeed, every well-defined formal system can be translated into a Turing Machine which enumerates all of the theorems derivable from its axioms. Conversely, one can select from the numbers enumerated by a Turing Machine those that are Gödel numbers of formulas in a given language. The deductive consequences of these formulas then result in a formal system's theorems. The second relevant equivalence maintains that, by way of an association between axioms and polynomials, the question concerning the consistency of a formal system's axioms can be reduced to a question in number theory. In particular, it can be reduced to the problem of determining whether a certain Diophantine equation has integer solutions for any possible integer values of its polynomials' parameters.

With these facts in mind, recall that Gödel's Second Incompleteness theorem states that for any formal system T, if T is consistent and contains a minimal amount of arithmetic, then the consistency of T's axioms cannot be proved by T. Hence, the question of T's consistency is undecidable in T. From this theorem, then, it follows that objective mathematics is inexhaustible, and that no finite rule could ever produce all of its evident axioms. Such is the case, Gödel argues, because the consistency of objective mathematics must be a true mathematical proposition, but which can never be proved by any formal system, according to his Second Incompleteness Theorem. Still, this fact this does not preclude the possibility of a well-defined system might contain all of subjective mathematics, for there might exist a finite rule which produces all of the evident axioms of subjective mathematics.

Gödel's argument for his initial disjunction then proceeds as follows: suppose that the human mind can be reduced to some Turing Machine which, in turn, corresponds to a well-defined formal system T, containing all of subjective mathematics. By Gödel's Second Incompleteness theorem, the consistency of T's axioms cannot be proved within T itself. Thus, the proposition concerning the consistency of T's axioms is entirely undecidable. Now, by the association between Diophantine problems and the question of consistency, it follows that there must exist absolutely unsolvable Diophantine problems, in the sense that no human mind – as a finite machine – could ever arrive at their answer. Thus, Gödel has argued that if the mind is a Turing Machine then there must exist absolutely unsolvable Diophantine problems. This statement, then, is logically equivalent to the statement that either mind is not a Turing Machine, or there exist absolutely unsolvable Diophantine problems, where the “or” in question is an inclusive one. That is, Gödel argues that if the human mind is confined to a Turing Machine, then there are certain mathematical problems which can never be answered, regardless of any possible advances within the field. Now, Gödel firmly believes that the human mind can know, with absolute certainty, that mathematics is consistent. Nonetheless, as has been established, such a proposition can never be mathematically proved. Gödel therefore believes that our minds possess a least one insight that can never be mathematically derived. Thus, the first disjunct of his statement can be rewritten as stating that the mind, if it is not a Turing Machine, “infinitely surpasses any finite machine”. Where the notion of a machine and a Turing Machine are taken as equivalent, by the Church-Turing thesis.

I now proceed to explain Gödel’s argument for the notion that, if the first disjunct in his statement in true, then the human mind cannot be fully captured by the workings of the brain. This notion follows naturally from Gödel’s belief that the human brain “to all appearances is a finite machine with a finite number of parts, namely the neurons and their connections”. That is, Gödel believes that the human mind can be reduced to a finite Turing Machine. Such is the case, presumably, because all of its trillions of neurons can be thought of as executing some simple instruction, as a Turing Machine does. This Machine’s tape would invariably have to be a finite, however, given that the human brain is limited in physical space. Hence, if the first disjunct of Gödel’s initial statement is taken to be true, such that the mind surpasses the powers of any finite Turing Machine, it then follows that the mind must also surpass the brain, if the brain is taken to be equivalent to some finite Turing Machine. Indeed, Gödel believes that the human mind can be fully aware of the consistency of mathematics, whereas the human brain – as a Turing Machine – could never derive such a fact. Still, though certainly a thought-provoking claim, Gödel’s conclusion that if the mind surpasses a Turing Machine, then its workings must surpass those of the brain rests explicitly on the assumption that the human brain can be reduced to a finite Turing Machine. It remains the case, however, that modern neuroscience can neither confirm nor reject this hypothesis. Still, I think there are reasons to be skeptical of this notion and that, at the very least, it is not as trivial as Gödel might think. I now proceed to discuss these thoughts.

On the one hand, it is clear that a human brain is at least as powerful as a Turing Machine, in the sense that it can be trained to mimic any computation that a finite Turing Machine can perform. Indeed, given a set of clear instructions, a person could easily sit down and recreate all of the simply steps executed by a Turing Machine – provided, of course, that physical limitations concerning time and space are idealized away. Thus, the important question to answer if one is interested in investigating the equivalence relation between Turing Machines and human brains is whether or not a human brain surpasses, in some sense, the powers of a finite Turing Machine. If the answer to this question is positive, then the brain should not be reducible to a finite Turing Machine. Now, one evident way of to demonstrate such notion would be to provide an example of a task that a human brain could perform, but that no finite Turing Machine could ever replicate. Sill, describing such a task is no simple matter, since Turing Machines are designed for the exact purpose of quantifying the mechanical processes undertaken by a human computer. It might be argued, then, that all human-executable tasks – from the most trivial arithmetic computations, to the most elaborate productions of fine art – can somehow be idealized into some algorithmic form, which can then be replicated by a finite Turing Machine.

Instead, I argue that there appear to be particularities concerning the design of a finite Turing Machine that might not suffice to capture the fundamental complexity of the human brain. For one thing, a defining characteristic of a Turing Machine is that it operates under discrete time. That is, a Turing Machine follows a set of instructions provided by a program while moving about a countable set of integers, each of which represents a discrete point in time. Nonetheless, it is not at all clear that the brain's computations proceed a similar manner. Indeed, it seems to me that it would be far more plausible to think of the brain's computations as operating over some continuous interval of time. For, to the extent that one can think of time as a continuous variable over the real number line, it seems hard to imagine that the brain would only operate over one of its countable subsets. Since the brain is always active, it seems to me that if one were to consider an interval of time of infinitesimal length, it could still be shown that the brain would be in the process of executing a multitude of different computations during this miniscule interval.

As a counterpoint, however, one might consider whether time itself can even be thought of as continuous. That is, while physicists do typically model time as continuous variable over the real number line, it is not at all clear that this is represents its true nature. Now, if one is willing to consider the possibility that time is truly discrete, one might further endeavor to contemplate the possibility that all of our physical reality is reducible to Turing Machines. If such a bold hypothesis were to be true, then it would certainly follow that the brain is reducible to a Turing Machine, given that the brain represents a finite subset of our physical reality. Nonetheless, there are perhaps even stronger reasons for believing that this hypothesis does not hold. For one thing, modern advances in physics seem to have moved away from a deterministic understanding of the processes at the quantum scale, and it may very well be the case that the such complex phenomena can only ever be formulated in genuinely stochastic terms. If this is indeed the case, and true randomness is inherent to our universe, then it would seem most unlikely that the archetypal Turing Machine – a strictly deterministic device – could ever account for all of reality.

Now, as to the question of the brain's mechanical nature, one might contemplate the idea that – in a similar fashion – the brain can only ever be understood in terms of genuinely random processes at the quantum scale. Indeed, proponents of theories of quantum consciousness have defended the idea that human consciousness might not be explainable in terms of classical physics, and instead must be explained at the level of quantum mechanics (Penrose, 1989). If this were to be case, then it may very well follow that there are fundamental processes within the brain that can never be replicated by a traditional Turing Machines. Hence, even if it is the human mind does indeed surpass every finite Turing Machine, it might still not surpass the human brain. To be certain, such a hypothesis lies from the reach of contemporary neuroscience. Still, given the truly unique nature of human consciousness, I not think it impossible that our brain's must operate at a non-deterministic scale of quantum mechanics. At any rate, it remains the case that very little is known for certain about the nature of human brain and its obscure relationship to human consciousness. Given its remarkably complex nature, then, it is not inconceivable to imagine that the brain is inherently more powerful than any finite Turing Machine. It seems to me, then, that the question of the brain's reducibility to a finite Turing Machine is far less trivial than Gödel might have thought. For this reason, I do not believe that if the mind somehow surpasses every finite Turing Machine then it must necessarily surpass the human brain.

In this paper I have sought to show how Gödel's complex initial disjunctive statement lies at the very core of his philosophy of mind. Indeed, Gödel seems to have believed that the first disjunct of his statement holds true, and that the human mind therefore cannot be reduced to the workings of the human brain. In my response, I have argued that even if the human mind cannot be reduced to a Turing Machine, it does not necessarily follow that it cannot be reduced to the workings of the human brain. For indeed the nature of the human brain might be so incredibly complex as to surpass that which can be replicated by a Turing Machine, in which case it may still very well be responsible for all the workings of the human mind.

References

Penrose, Roger (1989). *The Emperor's New Mind*. New York, NY: Penguin Books.