4.5
Why is Human Self-Consciousness Different from Artificial Intelligence and Animal Consciousness?

Robert J. Spitzer, S. J., *Magis Institute, Irvine, California, U. S. A.*

As the reader will soon discover, a plethora of methodologies is used in the forthcoming analysis—direct longitudinal comparative observation, neurophysiological analysis (studies of the brain and nervous system), phenomenological analysis (describing and assessing our own inner states of consciousness), logico-mathematical analysis (e.g., Gödel's Theorem), ontological analysis (assessing the condition necessary for the possibility of a particular phenomenon), and transcendental analysis (assessing the origins of our desires for perfect and unconditional truth, love, goodness, beauty, and being).

One might be thinking, "Is it really necessary to use so many different methods? Wouldn't it be better to just stick with one or two, and do them really well?" Valid as these questions are, they really do not reflect the complexity of the subject with which we are dealing—namely human intelligence and self-consciousness. I hope it will become intuitively obvious that the subject warrants six different methodological approaches, because no one approach can even begin to describe, let alone explain, the height, depth, and breadth of this most remarkable phenomenon. Indeed, as we shall see, using six different methods still falls far short of manifesting the full reality. Nevertheless, six methods is better than one or two, because there is strength in seeing both the overlapping and the distinct pieces.

John Henry Cardinal Newman's notion of "informal inference" can help us to see the benefit of multiple methodologies more deeply . He noted that multiple data sets (arising out of different methods and observations) that converge upon a single conclusion afford two benefits:

> (1) they corroborate each other, and (2) they complement each other.

Corroboration is extremely important, because all complex theories are subject to modification and change as new discoveries are made. Now, if a fundamental conclusion is based on the convergence of, say, ten data sets generated from six different methods, then the modification of one of those data sets is not likely

to significantly alter the fundamental conclusion. However, if the fundamental conclusion is based on, say, two data sets from one method, then a new discovery could (and probably would) significantly alter the fundamental conclusion. Every fundamental conclusion is strengthened by the convergence of additional data sets (with distinct methodologies) having their own distinct antecedent probability.

With respect to complementarity, it is clear that direct, longitudinal, comparative observation can go a long way in grounding a theory about human consciousness, but it cannot do everything that an analysis of the brain and nervous system can do (and vice versa). Similarly, both kinds of empirical analysis (direct observation and neurophysiological) cannot do everything that phenomenological analysis and ontological analysis can do (and vice versa). Even though there can be considerable overlapping (which leads to corroboration), there will always be additional data leading to further exploration which will enhance the breadth and depth of the conclusion.

For these reasons, this brief analysis of human, animal, and artificial intelligence will focus on the differing methodologies of brain physiologists (such as Carruthers, Lurz, Rosenthal, Schumaker, and Smith), the longitudinal comparative analyses of researchers (such as Piaget, Boehm, Stubblefield and Richard), the phenomenological analysis of Hubert Dreyfus' "unconscious fields" (developed by Merleau-Ponty), the logico-mathematical analyses of mathematicians and physicists (such as Gödel, Lucas, and Penrose), the ontological analyses of philosophers (such as Lonergan and Rahner), and the transcendental analyses of philosophers (such as Plato, Augustine, Newman, Lonergan, and Rahner). This study will not delve into the important details of each method and study (because it would make this project completely unwieldy), but rather, will *summarize some* of the important findings of each of the above methods to give the reader the opportunity to see the *convergence* of all of them. The citations made here to various researchers and philosophers will provide a virtually endless resource for future study.

This study is divided into two parts, because we are comparing human intelligence to two quite distinct other kinds of intelligence. These two parts are: 1) Why can't human self- intelligence be replicated by artificial intelligence? and 2) How is human intelligence distinct from animal intelligence?

I

WHY CAN'T HUMAN INTELLIGENCE BE REPLICATED BY ARTIFICIAL INTELLIGENCE?

Hubert Dreyfus presented one of the first systematic treatments of the limits of artificial intelligence. In his two classic works (*What Computers Can't Do* and *What Computers Still Can't Do*), Dreyfus, using concepts from Heidegger and Merleau-Ponty, showed that human intelligence depends primarily on unconscious fields rather than conscious symbolic manipulation, and that these unconscious fields cannot be captured in formal rules. This gives rise to the question concerning the nature

of these unconscious fields, and we will discuss this in detail below, in Section II. For the moment, we will want to investigate the dependence of artificial intelligence on formal rules, and the apparent ability of human beings to transcend such rules.

IA. Godel's Theorem

The famous mathematician Kurt Gödel gave a formal proof (Gödel's Theorem) that showed that the way humans do mathematics is not replicatable by machine intelligence (which is dependent on given rules and algorithms). It seems that human intelligence can always get beyond any algorithm, notice the algorithm's inadequacy compared to other possibilities, and then create something truly new (which is not related to the previous set of rules or algorithms) (See Satze).

Several physicists have refined Gödel's proof to meet contemporary challenges, and these proofs continue to have considerable probative force (See Lucas, Penrose, *Emperor,* and Barr). In brief, Gödel showed that there will always be unprovable propositions within any set of axiomatic statements in arithmetic. Human beings are able not only to show that consistent, unprovable statements exist, but also to prove that they are consistent by making recourse to axioms beyond those used to generate these statements. This reveals that human thinking is not based on a set of prescribed axioms, rules, or programs, and is, by nature, beyond any program.

A deeper explanation of Gödel's theorem may prove to be helpful. Stephen Barr, summing up the Lucas version of Gödel's argument, notes:

> First, imagine that someone shows me a computer program, P, that has built into it the ability to do simple arithmetic and logic. And imagine that I know this program to be consistent in its operations, and that I know all the rules by which it operates. Then, as proven by Gödel, I can find a statement in arithmetic that the program P cannot prove (or disprove) but which I, following Gödel's reasoning, can show to be a true statement of arithmetic. Call this statement G(P). This means that I have done something that that computer program cannot do. I can show that G(P) is a true statement, whereas the program P cannot do so using the rules built into it. ¶ Now, so far, this is no big deal. A programmer could easily add a few things to the program—more axioms or more rules of inference—so that in its modified form it can prove G(P). (The easiest thing to do would be simply to add G(P) itself to the program as a new axiom.) Let us call the new and improved program P′. Now P′ is able to prove the statement G(P), just as I can. At this point, however, we are dealing with a new and different program, P′, and not the old P. Consequently, assuming I know that P′ is still a consistent program, I can find a Gödel proposition for it. That is, I can find a statement, which we may call G(P′), that the program P′ can neither prove nor disprove, but which I can show to be a true statement of arithmetic. So, I am again ahead of the game. . . . This race could be continued forever (Barr 214).

Since human beings can indefinitely prove propositions which are not provable through the axioms from which they were derived, it would seem that human intelligence is indefinitely beyond any axiomatic or program-induced intellection.

Several attempts have been made to critique the strength of Lucas' claims concerning the non-mechanistic (non-formal and non-computational) dimensions of human thinking (See McGill). Lucas has responded to these criticisms by clarifying the terms and parameters of his original formulation of the Gödel proof. Though Lucas' critics point to ambiguities in his use of "consistency" and "inconsistency," and his use of idealizations within both artificial and human intelligence, these criticisms do not undermine the *general* validity of his rendition of Gödel's proof. Human intelligence appears to be beyond the mechanistic structures and processes of computational processing and formal rules (See Lucas).

IB. Penrose's Version of Godel's Theorem

In 1994 (*Shadows of the Mind*), Roger Penrose took a different approach to responding to critics of his 1989 formulation of the Gödel proof (in *The Emperor's New Mind*) by reformulating the proof. Essentially, Penrose shows that a contradiction will result if one assumes that one's reasoning powers (and one's awareness of the truth of one's reasoning powers) can be captured by any formal system (See also Chalmers). Therefore, human reasoning powers (and the soundness or truth of one's reasoning powers) cannot be captured by any formal system.

This version of the proof is stronger than the 1989 version, but has been criticized for its assumption that human reasoning powers are sound (true) and that humans are aware of the truth of their reasoning powers (Chalmers). This criticism focuses mainly on the fact that mathematicians make mistakes, and so they cannot have been aware of the soundness of their reasoning at the time the mistakes were made. Penrose responds by noting that mathematicians know certain unassailable truths (which can be shown to be unassailable), and that they distinguish these propositions from fallible (or correctible) propositions. Furthermore, if mistakes are made, mathematicians can recognize these mistakes and reveal their falsehood through the unassailable propositions. Thus, human beings (mathematicians) can know both the soundness or truth of the unassailable propositions and the soundness or truth of their reasoning when they grasp these unassailable propositions. However, artificial or computational intelligence (as modeled by a Turing Machine) cannot know either of these truths (otherwise a contradiction will result according to the Penrose version of the Gödel proof). (See Penrose, "Beyond" 23). [As scientists know, a "Turing Machine" is a hypothetical device described by Alan Turing in 1936 that models the basic mechanical, logical, and computational properties intrinsic to every computer. Turing's model has been sufficiently generalized to replicate any form of mechanical intelligence].

There is considerable agreement among mathematicians that they can know certain unassailable truths (and can therefore know the soundness of their reasoning based on these unassailable truths), and so the Penrose version of the Gödel proof should be taken quite seriously as a reasonable and responsible articulation of a radical difference between human intelligence and artificial intelligence. Human intelligence appears to be aware of mathematical generalizations that underlie the intelligibility and internal consistency of arithmetic and even

mathematics itself. This grasp of mathematical generalizations includes an awareness of their validity (and in some cases, their unassailable validity). In contrast to this, artificial intelligence (as modeled by universal Turing Machine) appears to have no awareness of such generalizations, and therefore, cannot have an awareness of the validity (or unassailable validity) of them.

Several theorists have presented other compelling proofs of the differences between mechanical and human intelligence. One of the more important proofs given by McCall shows that a Turing Machine must associate truth with provability while human intelligence can grasp truth independently of provability in many important respects (McCall).

In sum, we can say that the Gödel proof is alive and well. Not only are there multiple versions of the Gödel proof, but also many mathematicians and philosophers believe that they are capable of grasping the generalizations and validity of mathematics at a level that universal Turing Machines are incapable of doing. The original formulation of the Gödel proof, as well as the multiple reformulations of it by Lucas, Penrose, and others, indicate an intrinsic limit to universal Turing Machines that human beings seem to naturally transcend. This presents a significant challenge to the possibility of strong artificial intelligence (mechanical intelligence thinking at the same level as human intelligence).

In view of this chasm between artificial and human intelligence, we must ask ourselves three questions: (1) How does human intelligence come to have an awareness of generalizations underlying the intelligibility and internal consistency of arithmetic and mathematics? (2) How do humans know the truth (soundness or validity) of these arithmetical and mathematical generalizations? (3) How do humans know the unassailability (universal and necessary truth) of some arithmetical and mathematical generalizations (termed, "axioms")?

There are two contemporary theories which offer the possibility of an explanation: (1) The quantum explanation of Roger Penrose and, (2) The theory of heuristic anticipation offered by Plato, Polanyi, Whitehead, and Lonergan.

IC. Penrose's and Hameroff's Quartum Computation in Brain Micrtotubules

The validity of Penrose's version of Gödel's Theorem shows at the very least that human consciousness manifests a non-computational and non-algorithmic process which is not completely deterministic. This means that human consciousness must be doing something beyond the deterministic causal parameters of classical physics (because it goes beyond the computational and algorithmic properties of classical deterministic causal systems). Conversely, the human mind is not completely random—it clearly manifests orderly reductions to classical physical states; therefore, it cannot be described through quantum physics alone. This gives rise to a fundamental problem that requires an explanation if physicalist theories of human consciousness are to be maintained—namely, if human consciousness is capable of activities beyond the computational and algorithmic parameters of the deterministic causal systems of classical physics and it manifests an order which is not purely random (indicating reduction from a quantum to a

classical state). Then, is human consciousness beyond physical (materialistic) causal systems?

[I intentionally use the term "consciousness" here to designate a category which may go beyond the brain which is a physical-organic entity that would have to be explicable in terms of either classical or quantum physics. Human consciousness' capacity to grasp mathematical intelligibility, the validity of mathematical intelligibility, and the unassailability of mathematical axioms may well go beyond both classical (deterministic) physical systems and quantum (random) physical systems. If the proposals of Penrose and others to find a middle ground between classical and quantum physics prove unsatisfactory, it may well require that human consciousness be viewed as a reality beyond a merely physical or materialistic brain.]

Penrose and others (Benioff, Shor, Deutsch and Josza, and Hameroff) have attempted to find a middle ground between the deterministic systems of classical physics and the random systems of quantum physics in the human brain. These proposals attempt to find a domain of physics which can, as it were, have the best of both worlds—the non-computational and non- algorithmic dimensions of random quantum systems and the coherent ordered characteristics of deterministic classical systems. The dynamics of a quantum computer seem to hold out the best possibility for a physical explanation to this mystery of human consciousness. Quantum computation offers the possibility of moving beyond the strictly deterministic and algorithmic characteristic of classical computation through superposed states using multiple computations simultaneously, in parallel.

There are, however, two major obstacles to quantum computing which may prove to be vexing—interfacing input-output to the system and protecting the system from environmental decoherence. "Input-output interface with the system" refers to the challenge of integrating quantum inputs and outputs with classical ones. In order to achieve a successful interface between quantum systems and classical systems, it will be necessary to convert quantum inputs to classical ones, classical inputs to quantum ones, and quantum outputs to classical ones (in a way that maintains the essential character of both sets of inputs and outputs). This interface problem presents a multitude of challenges which may, in principle, be insurmountable because of the intrinsic limits of quantum and classical systems. "Environmental decoherence" refers to the problem that occurs when a very weak quantum input or output is made incoherent by the environmental noise of much larger and stronger classical systems with which the quantum system is interacting. Hameroff describes the problems as follows:

> At first glance the possibility of macroscopic quantum states in biological systems seems unlikely, appearing to require either extreme cold (to avoid thermal noise) or laser-like energetic pumping to achieve coherent states. And as in technological proposals, perfect isolation of the quantum state from the environment (and/or quantum error correction codes) would be required while the system must also somehow communicate with the external world (Hameroff, *New Frontier*).

Roger Penrose and Stuart Hameroff have forwarded a proposal which they believe *may* provide the beginnings of a solution to both problems—quantum computation in brain microtubules. In this proposal, protein assemblies called microtubules within the brain's neurons are viewed as self-organizing quantum computers. Penrose believes that he can avoid the problems of environmental decoherence and the problems of interfacing input- output to the system by appealing to a natural feature in space-time geometry. As Hameroff notes,

> Roger Penrose ("Beyond," *Emperor, Shadows*) has proposed that isolated quantum systems which avoid environmental decoherence will eventually reduce nonetheless due to an objective threshold ("objective reduction" - OR) related to an intrinsic feature of fundamental space-time geometry. Unlike the situation following environmental decoherence, outcome states which reduce due to Penrose's objective reduction are selected by a non-computable influence on the deterministic, pre-reduction quantum computation. Noncomputability implies a non-algorithmic process which is neither deterministic nor random, a property which Penrose also attributes to conscious thought and understanding. (See *The New Frontier*; also Hameroff and Penrose, "Conscious events").

So how does Penrose's theory work? By using the quantum gravitational dimension of space- time and reducing mass to the curvature of space-time, Penrose shows the *possibility* of an objective reduction from quantum to classical states that appears to avoid the two problems mentioned above (input-output interface to the system and environmental decoherence). He then applies this feature to microtubules in the brain (connecting the quantum system to the macroscopic neural system of the brain). A microtubule is a self-assembling hollow cylinder constituting the cytoskeleton (webs of protein polymers) which functionally organize cells. These cylinders *may* be sensitive to quantum activities (which are in turn subject to Penrose's objective reduction). If this theory can be verified, it would seem that Penrose has entered into the zone of "the best of both worlds," in which non-computational, non-algorithmic, activities can occur in nature (and in nature's brain).

Ingenious though this proposal is, it is very speculative, and up to now there has been no evidence that such objective reductions really occur in microtubules in the brain, or, for that matter, in nature. Furthermore, the proposal has been criticized on many objective grounds: The fundamental problem that physicists have with the Penrose proposal is summed up by Hameroff's question—how could near-infinitesimally small, weak and fast processes (in Penrose's objective reduction of quantum states) have macroscopic effects in biological systems (microtubules in the brain)? When one looks at the quantities involved in this proposal, this problem seems almost insurmountable. Tuszynski and Brown articulate the multiple problems with these quantities (and the physics underlying the proposal) which Hameroff summarizes as follows:

> The Planck length is 24 orders of magnitude smaller than the diameter of an atom. Approximately 10^{78} discrete Planck scale volumes correspond to the space occupied

> by one protein, and 10^{105} such volumes to a brain. The energy of one proposed Orch OR (e.g. 25 msec) is only 10^{-28} joules, or 10^{-10} electron volts (eV) whereas the energy of thermal noise (kT) is much larger at 10^{-4} eV. (*The New* Frontier).

In addition to these problems, Stanley Klein articulates two conjoint problems: (1)The brain operates at too high a temperature and is made of floppy material (not particularly sensitive to quantum effects); (2) Penrose has not explained how synaptic modulations can be achieved quantum mechanically, but not classically (See Klein).

These and other problems make the Penrose-Hameroff proposal unlikely. So where does this leave us? If the Penrose-Hameroff proposal proves to be physically unrealistic, then physicists will have to find another way of producing a naturalistic objective reduction from a quantum to a classical state which does not encounter the two fundamental problems of input- output interface with systems and environmental decoherence. Even if another explanation is found (besides microtubules in the brain), it is difficult to believe that there will not be extreme difficulties overcoming the quantitative differentials between quantum activity and macroscopic systems. Why? Because the criticisms of Tuszynski, Brown, and Klein do not apply to Penrose's and Hameroff's proposal alone, but rather to proposals that rely upon a causal relationship between quantum systems and macroscopic classical ones.

If such problems prove to be insurmountable, then what alternatives do we have? Physicists will have to find another way in which non-computational, non-algorithmic characteristics can occur in macroscopic classical systems. As noted above, this may not, in principle, be possible because of the intrinsic limits of physical systems. If that were the case, then artificial intelligence (as modeled by a Turing machine) would not be able to replicate the non-computational, non-algorithmic characteristics of human consciousness. Human consciousness would always be beyond anything that physical systems (with their intrinsic quantum and classical limitations) can achieve.

Failure to overcome these problems suggests another consequence—namely, how human consciousness can perform non-computational, non-algorithmic functions if it is considered to be merely a physical system (a brain which is reducible to physical systems alone). Consider the following: if we reduce human consciousness to a human brain, and then we reduce the human brain to physical systems alone, and if non-computational, non-algorithmic functions cannot be replicated by a physical system, then it would be impossible for human beings to do non- computational, non-algorithmic thinking functions—but, in point of fact, we do (as indicated by our grasp of the unassailable truth of certain universal mathematical propositions). Thus, the additional consequence of not overcoming the three problems of quantum computation is that human consciousness may be transphysical (and even demonstrably transphysical). As will be discussed below, there are four other strong indicators of the transphysicality of human consciousness: (1) the need for *heuristic notions* in the human grasp of mathematics and use

of language (see below, sections I.D. and II.A.); (2) the absoluteness of human self-consciousness (see below, section II.B.); (3) the human desire for perfect and unconditional love, goodness, beauty, and being/home (see below, section II.C.); and (4) the evidence of verifiable data from near-death experiences suggesting the survival of human consciousness after clinical bodily death (see below, section II.D.). As will be seen, these five indicators of transphysicality not only indicate specific characteristics which cannot be replicated by artificial intelligence (modeled by a Turing machine), but also characteristics which do not appear to be present in non-human, animal consciousness and intelligence (see below, section II).

ID. MENTAL MAPS, ORGANIZATIONAL SUPERSTRUCTURES, AND CATEGORIES

The definition of "heuristic notions" will become clear through the following consideration. Recall what Penrose asserted in his restatement of Goedel's proof—namely, that most mathematicians (and others) grasp the unassailable truth of certain universal mathematical propositions. This led him to search for a physical explanation of non-computational, non- algorithmic thinking functions (in the domain of quantum computing) to explain how human beings could transcend artificial intelligence (the functions of a Turing machine). As we saw, such physical explanations face significant challenges which may be insurmountable. But let's suppose, for a moment, that Penrose and Hameroff had been successful in demonstrating that quantum computation could take place in a physical system and in the human brain. Would they have fully answered the question of how mathematicians and others can grasp the unassailable truth of certain universal mathematical propositions? I think most philosophers of mathematics and language would agree that they had not. Why? Because non-computational, non-algorithmic thinking functions do not explain how human beings can have *universal* thoughts—that is, thoughts that are not linked to specific individual things (that is, thoughts that go beyond "picture-images"). Notice that most of the words we use represent ideas that do not make reference to any individual perceptions or pictures within our imaginations. There are a few words that do refer to ideas about empirical perceptions or pictures in our imagination. Proper names refer to specific individual objects or images. So also do words like "rabbit" or "plant" or "rain," etc. These words can refer to a sort of "generalized picture" in human consciousness or in animal consciousness (see below, section II.A). Once a chimp has experienced several dogs, for example, it can be taught that a specific sign (in sign language) refers to all such four-legged barking entities. The chimp can associate the sign with a group of perceptions.

Yet, words (or signs) that refer to these perceptual ideas based on groups of perceptions with an empirical similarity constitute a very small number of the words we understand or use. For example, all of our prepositions, questions,

conjunctions, grammatical concepts, causal concepts, and most of our mathematical concepts, have no direct reference to a perceptual image or a group of perceptual images. Please read the above three paragraphs and classify which words have perceptual referents and which ones do not. As you can see, the vast majority do not. How can these words have meaning for us if words refer only to perceptions, groups of perceptions, or picture-images? There must be some other way in which the vast majority of our words and ideas having meaning. The answer lies in the human capacity to link words to relation—relationships among perceptual images and relationships among relationships.

Human beings have the remarkable capacity, as Bernard Lonergan notes, to move beyond any merely empirical residue (individuality, concreteness, perceptual specificity, and space-time specificity) and into a domain of universality through relationships and even complex sets of relationships (Lonergan 51–52). This explains not only our capacity to generate words that have no perceptual referents, but also our capacity to do logic and mathematics, to develop scientific experiments, and to create metaphors, poems, and all forms of creative writing. Well-known philosopher of language, Noam Chomsky, noted that the most elementary indicator of an intelligence capable of moving beyond perceptual ideas is the capacity to understand syntactical differences (e.g., the difference between "dog bites man" and "man bites dog"). Even though the words in this sentence all have perceptual referents, their *syntactical use* (e.g., in the ordering within a sentence) already shows a level of meaning going beyond merely perceptual ideas.

How do human beings do it? How do we leave behind the individuality (the empirical residue) of perceptual ideas, and move into the realm of universal ideas which do not have any empirical referent at all? How do we enter into the realm of relationships among perceptions and then relationships among relationships? The answer, as we shall see, lies in heuristic notions (or categories), which are like large conceptual maps that provide the superstructure through which we can organize virtually every kind of relationship among perceptions, and every relationship among relationships.

But we are getting ahead of ourselves. It might be beneficial to go back to our mathematical starting point, namely, our grasp of the unassailable truth of certain universal mathematical propositions. We might begin with a very simple example of a universal mathematical proposition—the standard definition of a circle: a line forming a closed loop, every point on which is a fixed distance from a center point. This definition is considered universal because it can be applied to every circle. It is, as it were, a definition of a "perfect circle" or a "pure circle." How did we get to this definition? Bernard Lonergan gives one possible scenario—from reflection upon a cartwheel.

When a curious mathematically oriented person, seeking a definition of a circle, looks at a cartwheel, he might notice that there are three parts to it—the rim, the spokes, and the hub. He might then begin to abstract from the specific empirical qualities of that cartwheel (the specific kind of rim, spokes, and hub) and might then reduce the hub to a point, and then the spokes to radii

(two-dimensional lines extending from the center point), and the rim to a two-dimensional circle. Notice that as he does this, he is grasping the *relationships* that give rise to the above definition of a circle. He sees the relationship between the central point, the radii, and the circle—and when he does this, he can understand each part through his grasp of the relationships within the *whole*. Thus, he can understand the circle from the relationship between the center point and the radii extending from it; he can also understand the radii from the relationship between the circle and the central point; and he can even understand the center point from the relationship between the circle and the radii. Notice how the *whole* provides the context for seeing all the *parts* in their *relationship* to one another. When one grasps the relationships within the whole, one has the basis for moving beyond a merely empirical residue—one can get to relational ideas which go far beyond merely perceptual ones.

This raises an interesting issue. If we need a "whole" to understand the relationship among parts, where do we get "wholes" from? What kind of "wholes" are we using to organize the relationships among various perceptual ideas in order to get to our non-perceptual (universal) ideas? Are all "wholes" perceptual wholes—like the cartwheel and even geometrical figures?

Evidently, they are not. Geometry (and geometrical wholes) are quite special because they can be perceived, but other "wholes" (such as the equal sign of an algebraic equation) cannot be perceived. Indeed, if you go back to any previous page of this article, you will notice not only that the vast majority of the words you are reading are not individual perceptions or perceptual ideas (as noted above), but also that the whole which will allow their relational meaning to be organized could likewise not be individual perceptions or perceptual ideas.

It might be helpful to give an illustration at this point. Every reader understands the concept of location. Someone asks you, "Where are you?" and you might tell them, "At this address . . . ," or "In Orlando, Florida," or "In the United States," etc. But location cannot be taken too lightly just because we all understand what is meant by it. A second look will reveal that location is the understandability of relative positions (positions in relation to each other). Notice that the address of my house is meaningless without some understanding of other addresses and the relationship among them. Notice further that you really couldn't understand these relationships among spatial positions without some conceptual organizing superstructure in your mind—like a mental map. Even small children have mental maps. They see not only that their house is in a different position from their friend's house, but they begin to understand that their friend's house might be "up" or "down" the street. They might also understand that when they are looking up the street, their friend's house is to the right or to the left. Later, they begin to understand the idea of addresses, and how all of these addresses "fit together." And later, they understand notions of north, south, east, and west. Even though their parents have to explain what the word "north" refers to, they understand that it is a concept that can be applied not only to "north of my house," but to any other relative position—including the North Pole.

So how do we do this? How are we able to go beyond an animal's capacity to remember the roads upon which it was traveling, to the understandability of relative positions—to left and right, up and down, north and south, to addresses, longitudes, and latitudes? There must be some organizing "whole" (like the cartwheel) that provides a superstructure (like a mental map) upon which to situate positions in their relationship to one another. Notice that this is not memorizing the roads on which I traveled; it is a mental map capable of organizing every position in the world, and even within the solar system, and even within the universe itself. It is a mental map for the understandability of all relative positions.

Notice one more important characteristic about this remarkable organizing superstructure—this mental map. It has to be a "higher" idea than the relational ideas it is organizing. The mental map, as it were, has to be "larger" than all of the relational positions it is organizing. As you may suspect, there are many other "mental maps" (superstructures that organize relational ideas) beyond that of location. There would have to be an organizing superstructure for time (because times, like places, are relational positions). There are still other superstructures to organize other relational ideas—such as similarities and differences, and causes and effects.

This insight was recognized by generations of philosophers, such as Aristotle, Aquinas, Kant, and Bernard Lonergan—among many others. Each one of these philosophers understood that we could not move from the individuality of perceptual ideas to the universality of relational ideas without some superstructure ("whole") to organize the relations. They also recognized that if the relational ideas organized by this "whole" were not individual perceptions or perceptual ideas, then the whole which organizes them could likewise not be an individual perception or perceptual idea. These wholes are higher ideas—larger mental superstructures—than the relational ideas they organize. Hence, if what's being organized by the whole is not an individual perceptual idea, then neither could be the whole that organizes them.

Aristotle had an implicit awareness of these large conceptual organizing superstructures ("wholes") in his ten categories—substance, quantity, quality, relation, action, affection (passivity), place, time (date), position, and state. He viewed these categories as the *highest* genera that could organize all other predicates (relational concepts).

Aquinas related Aristotle's categories to his highest category—being (*esse*). He showed how each category can be explained in terms of its "possession of being" (either being in itself, being through another, or through privation of being) (See *Summa Theoloigca* I, Q. 84, Art. 7). As will be seen below, this insight holds out the possibility of unifying the whole network of categories and predicate in human cognition (see below, I.E.—Lonergan's cognitional theory).

Immanuel Kant, in the *Critique of Pure Reason*, also recognized the need for large superstructures through which our relational ideas could be organized. Utilizing Aristotle's term "categories," he shows that these organizational superstructures must be pure concepts of understanding (that is, they are not derived

from perceptions of the world or individual perceptual ideas). They are like innate ideas that the human mind uses to recognize intelligibility in phenomena (perceptual ideas) (*Critique* B129). As such, they are the preconditions for any intelligible experience. Kant recognized that these categories are the highest forms of relating ideas to one another, and shows how they are intrinsic to all human judgments. His categories are similar to Aristotle's, but he organizes them in a different way. The categories are:

Quantity
Unity
Plurality
Totality

Quality
Reality
Negation
Limitation

Relation
Inherence and Subsistence (substance and accident)
Causality and Dependence (cause and effect)
Community (reciprocity)

Modality
Possibility
Existence
Necessity

Evidently, Kant recognized the need for these large "mental maps" (organizational superstructures) allowing us to relate all perceptual ideas to one another (and even to relate relational ideas to one another). However, he did not go as far as Aquinas in relating all of these categories to a single highest category or superstructure (being).

IE. LONERGAN'S UNRESTRICTED DESIRE TO KNOW AND THE NOTION OF BEING

In *Insight: A Study of Human Understanding*, Bernard Lonergan synthesized all of the above insights in a single comprehensive cognitional theory. He recognized, with Aristotle, Aquinas, and Kant, that perceptual ideas (ideas grounded in picture-images or groups of picture- images) are quite distinct from relational ideas (which are derived from the relations of ideas to one another). He also recognized that relational ideas constituted the vast majority of ideas used in human thinking and discourse (as distinct from perceptual ideas which are

connected to individual things or images—pictures). Like Aristotle, he realized that these relational ideas had to be derived from some organizational superstructure that form the contexts through which ideas are related. However, he does not identify these organizational superstructures with either Aristotle's or Kant's categories (highest genera). Instead, he associates them with the eight major questions ("what?" "where?" "when?" "why?" "how?" "who?" "how much?" and "how frequently?").

So how does Lonergan replace the categories with the eight major questions? He noticed that questions, like the categories, are ways of representing our highest cognitive organizational structures. The questions function very much like the categories in that they form the context in which cognitional contents can be interrelated with one another.

For example, the question "Where?" might be viewed as a large cognitional map that allows various cognitional contents to be related to each other as geographical positions (here, near here, there, over there). See the more extensive example given in I.D above.

The question "When?" may be viewed as another organizational superstructure providing a context for cognitional content to be related to each other as temporal positions (earlier-later).

The question "What?" provides the context for relating things to one another through similarities and differences. It should be noted here that these similarities and differences give rise to all kinds of cognitional contents that have no corresponding individual picture-thought. For example, when a child abstracts the idea of "life" from the difference between a rock and, say, a bird (or a group of living things), he does not have a picture-thought of "life." The word "life" makes sense only in the context of a relationship, where the difference between lifeless objects and living objects can be isolated.

The question "Why?" gives rise to the organizational superstructure allowing the relationship between cause and effect.

"How much?" and "How frequently?" give rise to the organizational superstructures relating quantities in space (how much?) and time (how frequently?).

Each kind of question represents an organizational context which allows cognitional contents to be placed in relationship with one another to give rise to "ideas" which have no direct corresponding picture-image—ideas concerning geographical position, temporal position, similarities, differences, causes, effects, and even quantities, frequencies, and so forth.

Lonergan thought that there was no need to postulate the categories (e.g., the categories of Kant and Aristotle) as distinct from the eight questions. The organizational superstructures intrinsic to the intelligibility of the questions are sufficient. So where do these organizational superstructures come from? Following Kant and Aquinas, Lonergan implies that they must be innate (intrinsic to human beings before any thinking or understanding can take place, and not derived from perceptions of the world or picture-images). Recall from above that the questions (categories) have to be higher ideas than the ideas they are

organizing. This means that the questions (categories) cannot be a perceptual idea (individual picture-image) any more than the ideas they are organizing. But this presents a problem. If we need the questions (categories) to derive relational ideas from perceptual ideas, then we could not have derived the questions (categories) from perceptual ideas. We would already have to have them and use them in order to derive them from perceptual ideas. Now if the questions (categories) are not derived from perceptual ideas, then they cannot be derived from the world of sense data (or even from the images of sense data), in which case, they must be innate. Can these innate ideas be hardwired in brain circuitry? It is highly unlikely, because brain circuitry is individuated (grounded in what Lonergan calls "the empirical residue"), and the questions (categories) cannot be grounded in an individuated empirical residue. Thus, the questions (categories) would have to be both innate and transcendental (not derived from the sensory world or from the individuated circuitry of the brain). [The term "transcendental" was borrowed from Kant, who noted "I call all knowledge *Transcendental* if it is occupied not with objects, but with the way that we can possibly know objects even before we experience them " (*Critique* A12)].

Lonergan also borrows the grand unifying insight of Aquinas that all the categories are ultimately connected to being. Aquinas looks at this question from the vantage point of ontology or metaphysics (i.e., "all that exists"), while Lonergan looks at it from the vantage point of epistemology (i.e., "all that is to be known"). In this regard, he explores how our notion of being is the basis or ground of every question that we ask (and therefore is the underlying unity of all questioning and the different kinds of questions coming from it—"what?" "where?" etc.). So what does Lonergan mean by the "notion of being"? He describes it succinctly as follows:

> [T]he notion of being penetrates *all* cognitional contents. It is the supreme heuristic notion. *Prior* to every content, it is the notion of the *to-be-known* through that content. As each content emerges, the "to-be-known through that content" passes without residue into the "known through that content." Some blank in *universal anticipation* is filled in, not merely to end that element of anticipation, but also to make the filler a part of the anticipated. Hence, *prior* to all answers, the notion of being is the notion of the *totality* to be known through all answers. (*Insight* 380–381. Italics mine)

Two aspects of this complex passage are germane for our inquiry: (1) "some blank in universal anticipation," and (2) "the supreme heuristic notion."

Lonergan's notion of "universal anticipation" in the cognitional process was first put forward by Plato in the Dialogue *Meno*. In this work, Plato shows that every question presumes an anticipation of its solution. Without some "pre-understanding" of what is to be known, the knower would never "know what he or she does not know," and therefore would not ask a question. Essentially, what Plato is asking is, "How can I know what I do not know—which I need to know in order to advance my knowledge further?" In other words, in

order to ask a question, we must have some awareness of something we do not yet know, and if we do not yet know it, how can it be present to our minds so that we can be aware that we do not know it?

Plato conjectured that we must have some kind of anticipatory awareness, which is a pre- cognitional, pre-articulated, and pre-thematic awareness sufficient to reveal *that* I do not know something, and even *what* I do not know in a pre-articulated way. This satisfies the two requirements of our vexing problem—(1) that we have some awareness of what we do not know, sufficient to ask a question, and (2) that this awareness *not be* the thematic and articulated knowledge that we are seeking when we ask a question.

Notice that the human capacity to ask questions that are not already articulated and thematic points to how humans can be genuinely creative. This, in turn, distinguishes human intelligence from artificial intelligence, which does not truly ask questions, but acts according to instructions, rules, and algorithms that have been programmed into it. Even though these rules and algorithms can sometimes appear to be questions, they really are nothing more than rules and algorithms made to look like questions.

Now we must ask the question, where did we get this pre-articulated, pre-thematic awareness of "what I do not know"? Lonergan believes he has the solution—it must come from an innate, pre-articulated, pre-thematic awareness of "what is to be known." If I have a pre- thematic awareness of "what is to be known," I can recognize both *that* I do not know and even *what* I do not know, sufficient to ask a question.

Lonergan calls this "innate, pre-thematic, pre-articulated awareness of what is to be known" a "heuristic notion." "Heuristic" refers to cognitional content, but not thematic or articulated content. It is a kind of vague content sufficient for making a guess or having an intuition—a kind of cognitive premonition. Within the scope of scientific inquiry, Albert Einstein was well aware of this intuition that leads to creativity and discovery –

> The intellect has little to do on the road to discovery. There comes a leap in consciousness, call it Intuition or what you will, the solution comes to you and you don't know how or why.

Thomas Hosinski says that Michael Polanyi describes this pre-cognitive, pre-articulated awareness as "a foreknowledge or heuristic anticipation of the solution that cannot be explicitly stated." Alfred North Whitehead describes it as a "dim apprehension" or a "vague anticipation." The great physicist Sir Arthur Eddington expressed this insight in a very explicit way:

> Science can scarcely question [the] sanction [that there are regions of the human spirit untrammelled by the world of physics], for the pursuit of science springs from a striving which the mind is impelled to follow, a questioning that will not be suppressed. . . . [T]he light beckons ahead and the purpose surging in our nature responds. (327–328).

Eddington's "trans-physical light that beckons human consciousness beyond its current state of knowledge" closely resembles Lonergan's "supreme heuristic notion" and "universal anticipation" because it recognizes that inquiry (particularly scientific inquiry) must be aware *that* it does not know, and even, in some pre-articulate way, *what* it does not know. This pre- articulate awareness of what we do not know points in the direction we must look to find the solution—like a "light beckoning ahead."

If Plato, Einstein, Polanyi, Whitehead, Eddington, and Lonergan are correct, then human cognition is able to transcend the limits of set rules and algorithms through a pre-thematic, pre- articulated awareness of what is more completely intelligible than what it currently knows. This pre-articulated awareness of what is more completely intelligible points in the direction we must go to discover and articulate that more complete intelligibility.

Why does Lonergan call this "pre-articulated awareness of what is more completely intelligible than what is currently known," "*the supreme* heuristic notion" or "a blank in *universal* anticipation"? Because he recognizes that it not only makes possible intelligent questioning, but also *unrestricted* questioning—a questioning that will never cease until the complete set of correct answers to the complete set of questions is known.

Lonergan moves one step beyond Plato, Einstein, Polanyi, Whitehead, and Eddington, because he recognizes that what he (the knower) really desires is not just the answer to a question that happens to be on his mind, but also complete intelligibility—the complete set of correct answers to the complete set of questions. He notices that his drive to understand is not focused only on his situational awareness of a particular "to be known," but rather on everything that is to be known. Every time he asks a question, he is desiring to know everything about everything—not *just* the answer to a particular question. The *unrestricted* desire to know is part of every particular question being asked, and so the answer to every particular question fills only one "blank" in *universal* anticipation.

Thus, for Lonergan, the "pre-thematic, pre-articulate awareness of what is to be known beyond what is known" is really a "pre-thematic, pre-articulate awareness of *all* that is to be known," and his particular question at a particular moment is a single instance within a much larger heuristic horizon that will drive him endlessly to the only satisfactory goal—a knowledge of everything about everything (the complete set of correct answers to the complete set of questions). This awareness of the *universality* of the "to be known" is the only adequate explanation for unrestricted curiosity and the "pure unrestricted desire to know."

Recall that our awareness of the universality of the "to be known" is pre-thematic and pre- articulate—it is, as Lonergan calls it, a "notion." So how can we refer to this universal notion? Perhaps the best way is to call it an "awareness of the complete*ness* of intelligibility" rather than an awareness of complete intelligibility. This indicates an awareness of "what is yet to be known" without implying a knowledge of "all that is to be known."

In sum, the awareness of the completeness of intelligibility incites us to ask questions *continuously* and *endlessly*, until the "totality of all that is to be known" is achieved. Our awareness that we have not yet arrived at "all that is to be known," will cause us to ask another question, and will even point us in the direction of where and how to find the answer. For Lonergan, it would not be possible to desire a knowledge of everything about everything sufficient to move the knower continuously toward this goal unless he had this pre-thematic, pre- articulate awareness of everything that is to be known. How could we *always* know that we have not yet reached the complete set of correct answers to the complete set of questions every time we answer a particular question, unless we had some awareness of what the complete set of correct answers to the complete set of questions would be like? How could we have an awareness not only of what we do not know, but also where and how to find the answer to "what we do not know" whenever we answer a question? How could our desire to know everything about everything be present in every question asked and every answer given (sufficient to continuously ask more questions until we have achieved our goal) unless we had this pre- thematic awareness of "everything to be known"? Without the notion of being (our awareness of the completeness of all that is to be known) we would not be able to continuously and endlessly grasp the incompleteness of what we know (and even grasp continuously and endlessly how we are to find the answer to what we do not know).

To say that human beings have "pre-thematic, pre-articulate, anticipatory awareness of only *particular* "to be knowns" (in response to particular questions) is to completely underestimate the true nature of human cognition. We are not animated solely by situational curiosity, but by universal curiosity—and this universal curiosity is present in every question we ask—every time the desire to know manifests itself.

What could be the source of this remarkable capacity within human beings? What is sufficient to engender a desire to know everything about everything? Indeed, what could enable us to see the incompleteness in "what is to be known" *indefinitely* (until we get to "all that is to be known")? What could enable us to see the insufficiency in *every* answer or set of answers which is not "all that is to be known"? Lonergan believes that the only possible *source* of this universal awareness would have to be the *idea* of everything about everything (the *idea* of complete intelligibility—the idea of the complete set of correct answers to the complete set of questions. Yet such an idea can only occur in an unrestricted act of understanding (thinking) which human beings clearly do not have. So how can we have a pre-thematic, pre-articulate awareness of the content of an unrestricted act of understanding, if we do not have an unrestricted act of understanding? The content of this unrestricted act of understanding would have to, in some way, be present to our consciousness without being grasped or understood our consciousness. It would have to be present as a horizon or a backdrop to every act of cognition—standing, as it were, as a goal to be achieved—as a target toward which we must aim to satisfy the objective of our desire to know. This

makes Lonergan's cognitional theory transcendental beyond the Kantian sense (our awareness of "the way that we can possibly know objects even before we experience them"); for he is suggesting that a higher consciousness (indeed, a higher consciousness containing the highest achievable idea) is somehow present to our consciousness as a horizon of complete intelligibility. This will be taken up in the consideration of a transphysical dimension of human consciousness in Section III below.

Before discussing this more fully, we will want to answer two lingering questions that were raised above—namely: (1) how does Lonergan's notion of being explain all of the categories (organizational superstructures) needed to ask particular questions (such as "what?" "where?" "why?" etc.), and (2) how does Lonergan's notion of being help us to understand the commensurability (or incommensurability) between human and artificial intelligence?

So how does the notion of being (which is at the ground of all questioning) give rise to the eight *kinds* of questions (organizational superstructures of relational ideas)? Intrinsic to our awareness of the completeness of all that is to be known must be all the ways in which reality can be known. The former must contain the latter; otherwise, we would not be able to recognize the completeness of *all* that is to be known. How can one have an awareness of the completeness of all that is to be known without simultaneously having an awareness of the *ways* in which all that is to be known can be known? These "ways" are the eight kinds of questions—the eight ways in which all reality can be known. If one of them is missing—say, the question, "why?"—then we would not be aware of cause-effect relationships, and we would not be aware of the completeness of *all* that is to be known. Thus, it seems that intrinsic to our pre-thematic, pre- articulate awareness of all that is to be known, is a pre-thematic awareness of the ways in which what is to be known can be known. We might then call the eight kinds of questions (organizational superstructures of relational ideas) "subsidiary heuristic notions" (because they are derivative upon the supreme heuristic notion—the notion of being).

The one, all-encompassing, transcendental notion of being makes possible our capacity to move from the world of perceptual ideas (linked to individual, space-time images) to the world of relational ideas underlying predicates (objects, syntax, grammar, mathematics, logic, ontology, and all the concepts that result from them in the natural sciences, social sciences, liberal arts, fine arts, etc.)—and not only this, but also our awareness of the horizon of all that is to be known through all of these relational ideas and organizational superstructures. If we feel hesitant about following Lonergan to the *universal* transcendentality of our notion of being, we will be faced with the unanswered and unexplained datum of human consciousness—our desire to know all that is to be known, and our capacity to *always* recognize when we have not yet reached that goal, and where we must go and what we must do to take the next step to achieving it. If we agree with Lonergan that we possess this remarkable capacity of consciousness, then we will either have to affirm within ourselves the existence of this supreme heuristic

notion (universal anticipation), or find some other explanation that is equally adequate. In either case, it is difficult to see how we will be able to completely escape some dimension of universal transcendentality in human consciousness. We will revisit this theme in conjunction with other manifestations of universal transcendentality and transphysicality in Section III below.

IF. CAN ARTIFICIAL INTELLIGENCE REPLICATE HUMAN INTELLIGENCE?

If Lonergan's assessment of the notion of being truly reflects the most complete explanation of our consciousness' unrestricted desire to know and cognitional capacity, then it would seem that artificial intelligence will never be able to replicate human intelligence, because the notion of being (the supreme heuristic notion) would always have to be beyond any physical system. Physical systems are individuated (restricted to space-time coordinates) while the notion of being is universal and comprehensive and therefore, cannot be individuated. Before explaining this, it is important to discuss a common misconception about the capacity of artificial intelligence.

It may be thought that artificial intelligence can do mathematics—after all, a Turing machine can add, subtract, etc. Furthermore, Turing machines can do logical operations, and can even be responsive to syntax and grammar. How does this differ from human intelligence—with its questions (awareness of the incompleteness of what is known)?

This takes us back to the question we were addressing at the end of section I.B, namely, will artificial intelligence ever be able to do non-algorithmic thinking (that is, not be dependent on a given set of algorithms, but rather, be able to reach beyond those algorithms to justify and correct them)? The above analysis can give us a clue to the answer. Recall Penrose's formulation of the Goedel theorem, which indicates that human beings are able to justify and correct any set of *known* algorithms because we are aware of certain "unassailable truths of mathematics." These unassailable truths allow us to know the validity of our mathematical reasoning, which, in turn, enable us to justify and correct any algorithm upon which our thinking might be based.

This sounds remarkably similar to Lonergan's cognitional theory discussed above. Notice that Penrose's and Lonergan's theories contain a similar characteristic linked to the human capacity to go beyond algorithmic thinking. Penrose's "unassailable mathematical truths" presume a knowledge of mathematical relations on the *highest* possible level—that is, a level which reaches down and touches every algorithmic proposition. As we probe this insight more deeply, we can detect the condition necessary for this knowledge—namely, Lonergan's notion of being.

Penrose's assertion that mathematicians can understand the unassailability of these truths (that is, the universal correctness of these truths) implies that we have at least a pre-thematic, pre-articulate awareness of the completeness of

mathematical intelligibility—the completeness of *all* that is to be known in mathematics. How could we know that a truth must be correct for all particular mathematical propositions unless we had some kind of pre-thematic grasp of the *whole* domain to which it could apply? So, what is this "whole domain to which a mathematical truth could apply"? It would seem to be a derivative of Lonergan's notion of being (the pre-thematic awareness of the completeness of *all* that is to be known). Our pre-thematic awareness of the completeness of all that is to be known would have to contain a pre-thematic awareness of the completeness of all that is to be known in mathematics.

The solution to the vexing problem of Göedel's theorem lies in discovering the conditions necessary for the possibility of being aware of "certain unassailable mathematical truths" capable of justifying the validity of mathematical reasoning, and in turn justifying and correcting any set of algorithms. Lonergan's notion of being seems to provide such an explanation by showing how human consciousness can be pre-thematically aware of the *whole* (completeness) of what is to be known in mathematics. This pre-thematic awareness would enable us to justify and correct *any* and *all* mathematical and algorithmic truths grounded in it.

Can a Turing machine which is based on a set of binary operations and fixed algorithms ever imitate the prethematic awareness of mathematical intelligibility? The answer is decidedly "no," because the awareness of mathematical intelligibility would have to be beyond any specific mathematical truth (or set of truths) that can be processed and ordered by binary operations and fixed algorithms. As such, the human grasp of mathematical intelligibility (manifested in the awareness of the unassailability of fundamental mathematical truths) would always have to be beyond artificial or machine intelligence.

It should be remembered that human beings are not in control of their prethematic awareness of the completeness of intelligibility (which includes our prethematic awareness of mathematical intelligibility). We are aware of this universally transcendental idea only as a horizon (or backdrop) of our acts of questioning and understanding, but we really do not understand (or grasp) complete intelligibility. That would require an unrestricted act of understanding, which we clearly do not have (since we continue to ask questions). This leads to the conclusion that our awareness of the completeness of intelligibility is given to us—that is, it is made present to us by the unrestricted act of understanding which is the source of the idea of complete intelligibility. We are, as it were, beholden to an unrestricted act of understanding making itself present to us as the horizon of our understanding. Though we use this horizon to continuously answer questions and creatively discover new truths in mathematics, science, social sciences, and the humanities, we do not ultimately control what we are allowed to use. We are active in our use of the horizon in every act of questioning and answering, but we are ultimately passive recipients of the horizon of complete intelligibility within us. Although we are individually actively creative and intelligent, we do so through a universally transcendent horizon which is not our own, and is not ultimately controlled by us. This will be taken up again in Section III.

WORKS CITED

Barr, Stephen M. *Modern Physics and Ancient Faith.* Notre Dame, IN: Notre Dame University Press, 2003.

Chalmers, D. J. "Minds, Machines, and Mathematics." *Psyche 2* (1996): 11–20.

Dreyfus, Hubert. *What Computers Can't Do*. New York: MIT Press, 1979.

—. *What Computers Still Can't Do.* New York: MIT Press, 1992.

Eddington, Sir Arthur. *The Nature of the Physical World.* London: Macmillan, 1928.

Godel, Kurt. "Uber formal unetscheidbare Satze der Principia Mathematica und verwanter Systeme I." *Nonatshefte fur Mathematik und Physik* 38 (191): 173–198.

Hameroff, Stuart. "The New Frontier in Brain Mind Science" in *Philosophical Transactions* Royal Society London (A) (1998): 1869–1896.

Hameroff, Stuart and Roger Penrose. "Orchestrated Reduction of Quantum Coherence in Brain Microtubules: A Model for Consciousness," in *Toward a Science of Consciousness, ed. S. R. Hameroff, A. Kaszniak and A. C. Scott.* Cambridge, MA: MIT Press, 1996.

—. "Conscious events as orchestrated spacetime selections." *Journal of Consciousness Studies* 3(1) (1996b); 3653.

Hosinski, Thomas. *Process, Insight, and Empirical Method: An Argument for the Compatibilities of the Philosophies of Alfred North Whitehead and Bernard J. F. Lonergan and its Implications for Foundational Theology.* Dissertaion submitted to the University of Chicago Divinity School for PhD. Web.

Klein, Stanley A. "Quantum Mechanics Relevant to Understanding Consciousness." *Psyche* 2 (1965): 3ff.

Lonergan, Bernard. *Insight: A Study of Human Understanding.* Toronto: University of Toronto Press, 1992.

Lucas, John R. "Minds, Machines, and Godel." *Philosophy* 36 (1961): 21.

McCullough, D. "Can Humans Escape Godel?" *Psyche* 2 (1996): 57–65.

McCall, S. "Can a Turning Machine Know that the Godel Sentence is True?" *Journal of Philosophy* 96 (1999); 523–32.

McGill, Jason. "The Lucas-Penrose Argument ab out Godel's Theoren." *Internet Encyclopedia of Philosophy.* Web.

Newman, John Henry. *Grammar of Assent.* Notre Dame, IN: Notre Dame University Press, 1978.

Penrose, Roger. "Beyond the Doubting of a Shadow." *Psyche* 2 (1996): 23ff.

—. *The Emperor's New Mind.* Oxford: Oxford University Press, 1989.

—. *Shadows of the Mind.* Oxford: Oxford University Press, 1994.

Penrose, Roger, and Stuart Hameroff, "What gaps? Reply to Grush and Churchland. *Journal of Consciousness Studies* 2(2) (1998): 99112.

—. "Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness." In *Toward a Science of Consciousness: the First Tucson Discussions and Debates, ed.* by S. R. Hameroff, A. Kaszniak and A. C. Scott. Cambridge, MA: MIT Press, 1996.

Shor, P. W. "Polynomial Time Alogrithms for Discrete Logarithms and Factoring on a Quantum Computer." In *Algorithmic Number Theory.* First International Symposium, ANTS 1 Proceedings, eds. L. M. Adleman and M. D. Huang. Berlin: Springer Verlag, 1994.

Tuszynski, J. A., and J. A. Brown. "Dielectric polarization, electrical conduction, information processing and quantum computation in microtubules, are they plausible?" New Frontiers, 1998.