Cosmos and History: The Journal of Natural and Social Philosophy, vol. 14, no. 1, 2018

ON QUANTUM MECHANICAL AUTOMATA, GÖDEL NUMBERS, AND SELF-REFERRING CONSCIOUSNESS[1]

Fred Alan Wolf

CONTENTS

[1] Editor's note: *Foundations of Mind,* the independent research group that has provided the papers for this special edition, has never taken either corporate or state money and is financed entirely by donations. Authors keep copyright without paying. The typical fee for this charged by open-access journals such as those published by PLOS, is around $2k. If you value this project, and wish to see further such proceedings from this group, we ask you to consider donating to *Foundations of Mind* – as little as $5 per download, through their website: http://www.foundationsofmind.org/donate. This will ensure there will be further published proceedings on the foundations of mind like this one for you and others to enjoy free.

ABSTRACT*:* In this paper I shall show how a post-quantum physical model of the self-referring mind based on Albert's quantum automata and Yurov's Gödelzing model of the same can be constructed. A quantum automaton seems to exhibit privacy of consciousness—the realization of an individual mind and what differentiates a "self-mind" from an "other-mind". My model, extending from Albert and Yurov, is based on consideration of qubits acting as quantum state vectors and quantum mechanical operators acting as quantum computer gates. I hope to show:

- The relationship of Gödel's undecidibility proof to David Z. Albert's model of quantum automata can be understood in quantum computing language.
- Whether or not we need new physics to understand self-referring quantum automata.
- The privacy of consciousness—why we each have the realization of an individual mind and what differentiates a "self-mind" from an "other-mind"
- The connection between objective and subjective experience.

In this rather simple manner, I believe I have explained how it is that our experiences of the "out there" world, that we know must include our memories in order to be perceived, appear to us as occurring "out there" even though we strongly suspect from neurophysiology that such experiences must be projected from our brains and nervous systems in some yet to be determined manner.

KEYWORDS: Quantum Mechanical Automata; Gödel numbers; Self Referring Consciousness

INTRODUCTION

For a period of time now I have been engaged in an ongoing effort to understand the everyday experience we all have called consciousness and its relation with the physical world.[2] Indubitably quantum physics enters into my research. A while ago I read David Z. Albert's work on quantum automata.[3] It then seemed to me to point in the direction

---

[2] My work will be found in the following sources:
Star Wave: Mind, Consciousness, and Quantum Physics. New York: Macmillan, 1984.
"The Quantum Physics of Consciousness: Towards a New Psychology," *Integrative Psychology* 3 (1985): 236-47.
The Body Quantum: The New Physics of Body, Mind, and Health. New York: Macmillan, 1986.
"The Physics of Dream Consciousness: Is the Lucid Dream a Parallel Universe?" *Lucidity Letter* 6, no. 2 (December 1987): 130-35.
Parallel Universes: The Search for Other Worlds. New York: Simon & Schuster, 1989.
"On the Quantum Physical Theory of Subjective Antedating." In *Journal of Theoretical Biology* 136 (1989): 13-19.
The Eagle's Quest: A Physicist's Search for Truth in the Heart of the Shamanic World. New York: Summit, 1991.
"The Dreaming Universe." *Gnosis* 22 (Winter 1992): 30-35.
"The Body in Mind." *Psychological Perspectives*: A Journal of Global Consciousness Integrating Psyche, Soul and Nature 30 (Fall-Winter 1994): 22-35.
The Dreaming Universe: A Mind-expanding Journey into the Realm Where Psyche and Physics Meet. New York: Simon & Schuster, 1994. Reprint, New York: Touchstone, 1995.
"The Quantum Mechanics of Dreams and the Emergence of Self-Awareness." In *Toward a Scientific Basis for Consciousness*, edited by S. R. Hameroff, A. W. Kaszniak, and A. C. Scott. Boston, MA: MIT Press, 1996.
"The Soul and Quantum Physics." In *Experiencing the Soul: Before Birth, During Life, After Death*, edited by Eliot Jay Rosen. Carlsbad, CA: Hay House, 1998: 245–52.
"The Timing of Conscious Experience." In *Journal of Scientific Exploration* 12, no. 4 (Winter 1998): 511–42.
"A Quantum Physics Model of the Timing of Conscious Experience." In *Toward a Science of Consciousness III*, edited by Stuart Hameroff, Al Kaszniak, and David Chalmers. Cambridge, MA: MIT Press, 1999: 359–66.
"The Quantum Physical Communication Between the Self and the Soul." In *Noetic Journal* 2, no. 2 (April 1999).
The Spiritual Universe: One Physicist's Vision of Spirit, Soul, Matter, and Self. Portsmouth, NH: Moment Point Press, 1999. Originally published as *The Spiritual Universe: How Quantum Physics Proves the Existence of the Soul*. New York: Simon & Schuster, 1996.
Mind into Matter: A New Alchemy of Science and Spirit. Portsmouth, NH: Moment Point Press, 2001.
Matter into Feeling: A New Alchemy of Science and Spirit. Portsmouth, NH: Moment Point Press, 2002.

[3] Albert, David Z. "How to Take a Photograph of Another Everett World." in *New Techniques and Ideas in Quantum Measurement Theory*. ed. D. M. Greenberger in Vol. 480. *Annals of the New York Academy of Sciences*. December 30, 1986. Also see: "On Quantum-Mechanical Automata," *Physics Letters*. 98A, no. 5, 6 (October 24, 1983), pp. 249-252, "A Quantum-Mechanical Automaton." *Philosophy of Science*. 54. No. 4

we need to take in science to understand a quantum physical model of mind. With the advent of the modern computer age, particularly with the advances in artificial intelligence (AI) and the rapid growth in building quantum computers, that new direction pointed to by Albert may be yielding a new understanding of consciousness—at least as far as we can provide new models.

It still seems to me that quantum physics and consciousness must be intimately related as a number of physicists and consciousness researchers have indicated over the past years.[4] Quantum physics indicates that an observer plays a crucial role in determining the objective qualities of the observable physical world. There does not seem to be any way to dismiss this role as a mere consequence of matter. Yet a key insight that Albert uncovered was the distinction between subjective and objective memory states that can arise within quantum automata.

Once a model of the mind appears promising, the next step would be to extend the model, although still keeping to the rules, rigor, and logic found in it. One might think that such an extension would be unpromising and, being based on quantum physics, would leave out far more than it could encompass. For example, the gap between purely subjective and objective experiences appears to be insurmountable. How can we expect the inner subjective world of the mind to be subject to the same laws as the outer objective world of matter? Why should we even have such an expectation?

In this paper I offer an extension of Albert's work on quantum automata by showing how quantum automata could function within the rules of quantum computation by consideration of qubits acting as quantum state vectors and quantum mechanical operators acting as quantum computer gates. I have discovered how such a memory could be constructed based on the language of quantum computation—namely through the use of qubits acting as quantum state vectors and quantum gates operating as quantum physical operators. Most import in this extension is the role played by eigenvalues being represented by metamathematical statements.

THE PROFESSOR AND HIS FRIEND

Let's begin with an old story from the early days of quantum physics and its dealings

(Dec. 1987), pp. 577-585, and *Quantum Mechanics and Experience*. Harvard University Press. 1992. pp. 180-189.

[4] There are many engaged in this endeavor as can be seen by looking through the literature. See, for example: Bass, Ludvik. "The Mind of Wigner's Friend," *Hermathena: A Dublin University Review*. No. 112, (1971), p. 58. Goswami, Amit. *The Self-Aware Universe: how consciousness creates the material world.* New York: Tarcher/Putnam, 1993. *Toward a Science of Consciousness III.* Edited by Stuart Hameroff, Al Kaszniak, and David Chalmers. Cambridge, MA: MIT Press, 1999.

with paradoxes. This one deals with the paradox caused when one tries to put the actions of an observer into the quantum physics picture.

In Fig. 1 we see a friend of Prof. Wigner carrying out an experiment involving a particle placed in a closed box. According to the usual understanding of quantum physics the bound-in particle no longer has a well-defined position, but now assumes the form of a standing-wave pattern inside the box. This pattern tells us where the particle is likely to be found but not where it actually is. Furthermore because the particle is contained, its momentum is also indeterminant—it could be moving either toward the left side or the right side of the box.

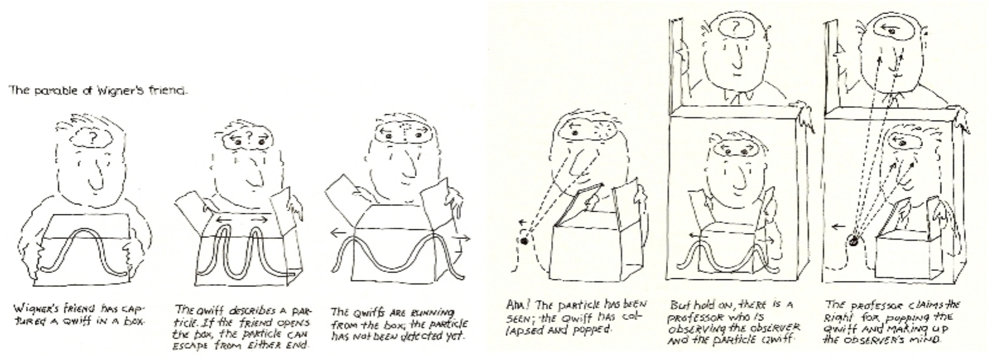

Figure 1. The parable of Wigner's friend (from Taking the Quantum Leap)[5]

To find out what's happening in the box, the friend opens the two opposite sides of the box simultaneously. The removal of the two sides together causes the wave pattern to split into two oppositely moving wave pulses. Both pulses pass out of the box, but then the friend "sees" the particle on his right side of the box and records his observation in his memory.

But unbeknownst to the friend his professor was observing the whole experiment of the particle, box, and friend, who wasn't aware of the Prof.'s presence. The professor explains to his friend that he was carrying out a curious experiment of his own that involved the particle, box, and friend, all placed inside a larger box. Following the rules of quantum physics, even the friend's observation of the particle, along with the wave, were split into possible editions with one edition having the friend seeing the particle on the left of the box and the other edition having the friend seeing the particle on the right of the opened box. The professor points out that it was his kind observation of the friend and the particle that 'created" the friend observing the particle when he (the

[5] Wolf, Fred Alan. *Taking the Quantum Leap*. San Francisco: Harper & Row, 1981. Revised Edition, New York: HarperCollins, 1989 pp 216-217.

Prof.) had opened his larger box. The friend and the particle owe their very existence to the Prof.'s kind observation.

So to whom should we give the honor of "creation"? What the parable exhibits is the difficulty of dealing with the role of observation and observers in quantum physics when one tries to put the so-called "collapse of the wave function" into quantum mechanics. Where, when, and who should we place the honor? There still does not appear to be an adequate answer to this multiple thronged question.

SEEKING AN ANSWER

In our world today the computer has become a tool that almost everyone on the planet has used or will learn to use in the very near future. Current sophistication of these machines allows people to obtain answers to problems by simply typing in their questions and waiting for answers that come without any human influence. Consequently the use of artificial intelligence (AI) has become a very active area of research. As far as I am able to discern all AI devices, as clever as they seem to be, operate according algorithmic procedures—lines of code that direct bits to change or not according to the input of instructions that are themselves lines of code. We can view such operations as constituting a formal system based on logical arithmetical axioms that we can label as W_c (the "c" standing for "classical" and the "W" for "World"). Hence computers, although they use quantum mechanics in order to work, operate on classical pieces of information called bits that can be written numerically as 0 or 1. We can think of these bits—long lines of them—stored in memory units we can call automata.

The question arises can these machines answer all questions? Can they exhibit intelligence in the same way that a smart human can do? Or is such a hope doomed? Many AI enthusiasts believe it is only a matter of time before AI computing devices will "take over the world." Let's look into this question. To do so I will need to corral our thinking somewhat. So let me assume we seek answers based on a W_c—classical formal system dealing with or based on arithmetic.

In his book,[6] Sir Roger Penrose discussed the amusing example of a chess match in which black has a decided advantage in material having two rooks and a bishop over white, besides all of its pawns, while white clearly is at a disadvantage with no pieces. Playing white, however, actually can force a draw by simply moving its king around behind its pawn fortress. Black's rooks and bishop are simply trapped behind the wall of black pawns opposing white's pawns. An AI machine such as IBM's deep thought was

[6] Penrose, R. *Shadows of the Mind.* New York: Oxford University Press, 1994. p. 46.

put into this game as white and asked to make a move. What did it do? It took the rook, thereby opening the wall to an eventual loss due to black material superiority. Clearly this was stupid, but logical given that an AI machine is preconditioned to win one move after another with material gain or superior position to be made upon each move. Yet the strategy implied by any novice chess player playing white is quite simple—play for a draw by just moving your king to any square not being threatened by a black pawn.

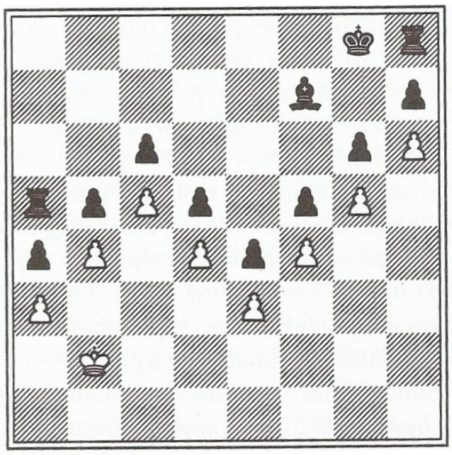

Fig. 2. What should white do to not lose the game? (Taken from Penrose).

PART I: A PATH FORWARD.

Since we ultimately are to compare machine intelligence with our own minds, it is natural to ask how do our minds operate. To answer we must enter the mysterious realm called consciousness. Let's first pose some more questions and possible paths forward:

- Can a search for a proper model of conscious experience be at first narrowed to the realm of the mathematical structures found in quantum physics?
- Should we expect the inner world of the mind to be like the outer world of matter?

To respond, it will be necessary to explore what is meant by knowing something and to do that we need to explore what we mean by "meaning." I shall use the term *meaning* in a certain specified way: Meaning occurs when a metamathematical statement is made about a mathematical statement. That's it! No hidden meanings. We may wonder at this point how such a simple definition can accomplish very much as we journey forward. I'll give a hint at this point: Gödel's inconsistency and incompleteness proofs.

META-VALUES OF OBSERVABLES

Before we get into Gödel let me illustrate what I mean by meta-statements.[7] Take the number 5. Compare it with the sign for 5 as written in Roman numerals, V, or in Hebrew as, h, for examples. In English I can write a meta-statement for 5 using single quotation marks as '5' to stand for the numeral for 5. It says in essence '5' is a sign designating the number 5. This use of quotation marks provides a meta-statement about the number 5.

In quantum mechanics we deal with concepts called quantum wave functions or quantum state vectors. These are abstract ideas in much the same way that numbers are abstract ideas. If I ask you to show me a 5, you draw a blank. You may ask 5? What is that? I can show you 5 fingers, or five toes, e.g., but 5 in and of itself, is abstract.

Suppose we have a quantum state vector representing the quantity five of something and just as we use the sign '5' to denote the number 5, we use a notation to denote the quantum state vector for state 5 as |5>. This kind of quantum state vector is called a *ket*.[8]

We also, following the work of John von Neumann, denote what happens to a ket when a measurement of the physical observable—carried out by a measuring instrument or automaton—which the quantum state vector refers, is carried out. In order to be read at a later date, the instrument or automaton must contain a memory record. When the measurement is completed the original quantum state vector is multiplied by another quantum state vector that represents the action of the automaton and refers to the value actually measured. The measured state that is contained in a memory device or quantum automaton is denoted $|'5'>_m$ with the sub-index 'm' denoting the operation used to make the measurement. It marks 5 of something, not just 5 itself. My point here is that a quantum automaton does not contain the number 5 any more than the word 'Chicago' contains the city although 'Chicago' contains 7 letters.

Hence in classical mechanics or using a classical logical system, denoted by W_c, to

---

[7] For more on this see Nagel, Ernest and James R. Newman. *Gödel's Proof.* NY: New York University Press. 1958. pp. 26-36.

[8] This is called the Dirac notation. See any standard book explaining quantum mechanics.

perform a measurement of some objective property, such as how many fingers I am holding up, is a question asking what is held in memory denoting a number, say 5. For a binary system with 3 bits 5 could be written '101' in the memory of the automaton. Furthermore quantum automata can carry out calculations—they can solve equations such as Schrödinger's equation of quantum physics. Given a metavalue of some observable and a procedure for doing arithmetic the quantum automaton can compute an answer, such as an average value or a predicted value of some observable. In brief they can interpret and show a prediction or answer to particular kinds of questions. When such a quantum automaton carries out such a procedure we say a measurement has been made upon some objective system.

In quantum mechanics performing a measurement of some objective property such as the direction of the axis of a spinning particle is a question asking what the quantum automaton holds in memory denoting the measured direction of that spin. To ask about such a memory means making an inquiry of a quantum automaton.

QUANTUM AUTOMATA QUBIT OPERATIONS

What makes quantum automata different from classical automata? The big difference is the kind of questions they can respond to, answer, and make valid predictions about. As I shall show, they are able to determine by scanning their own memories predictions about themselves that are in violation of the uncertainty principle.

Here we will explore such predictions using signs—metavalues—denoting quantum state vector measurements and what happens to these quantum state vectors when a measurement is completed. To keep thing as simple as I can here, I shall use quantum computer language to represent various states. I shall only consider 4 (binary) qubits or base states and their superpositions; these states (also referred to as the computational basis) are $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$. I'll refer to these states in two directions, along the z-axis, along the x-axis, and along both axes where one quantum state vector refers to x and the other to z. Thus we can write $|00\rangle$ as $|0_z0_z\rangle$, $|0_x0_x\rangle$, or $|0_z0_x\rangle$.

We also have a set of operators for which these quantum state vectors are eigenvectors. These operators are quantum gates through which we can envision the quantum state vectors flowing. We shall consider three types of gates: ZZ, ZX, and XX and their corresponding quantum state vectors that pass through them.[9]

[9] For those familiar with quantum mechanics, $Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. A term such as ZX means a tensor multiplication of the two separate operations, hence $ZX = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. You can look

For the computational basis, using $|0_z0_z>$, $|0_z1_z>$, $|1_z0_z>$, and $|1_z1_z>$we have using ZZ in eqns. 1A, 1B, 1C, and 1D,[10]

$$ZZ\ |0_z0_z> = `0_z0_z'\ |0_z0_z> \qquad \text{eqn. 1A.}$$

$$ZZ\ |0_z1_z> = `0_z1_z'\ |0_z1_z> \qquad \text{eqn. 1B.}$$

$$ZZ\ |1_z0_z> = `1_z0_z'\ |1_z0_z> \qquad \text{eqn. 1C.}$$

$$ZZ\ |1_z1_z> = `1_z1_z'\ |1_z1_z> \qquad \text{eqn. 1D.}$$

Appropriate linear superpositions of these states can also be written.[11] We get, e.g.,

$$(|0_z0_z> + |0_z1_z>)/\sqrt{2} = |0_z0_x>, \qquad \text{eqn. 1E.}$$

$$(|1_z0_z> + |1_z1_z>)/\sqrt{2} = |1_z0_x>. \qquad \text{eqn. 1F.}$$

Similarly we can apply ZX in eqns. 1E and 1F. The results for the above quantum state vectors are (since these are also eigenvectors of ZX):

$$ZX\ |0_z0_x> = `0_z0_x'\ |0_z0_x>, \qquad \text{eqn. 1G.}$$

$$ZX|1_z0_x> = `1_z0_x'\ |1_z0_x>, \qquad \text{eqn. 1H.}$$

resp.

Another appropriate linear superposition of these states 1E and 1F can also be written.[12] We get,

---

them up in Nielsen, Michael A. and Isaac L. Chuang, *Quantum Computation and Quantum Information*. NY: Cambridge University Press, 2014, pp 16-20.

[10] There are two points to be made here. First of all: $|0_z0_z> = \begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}1\\0\end{pmatrix}$, $|0_z1_z> = \begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}0\\1\end{pmatrix}$, $|1_z0_z> = \begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}0\\1\end{pmatrix}$, and $|1_z1_z> = \begin{pmatrix}0\\1\end{pmatrix}\otimes\begin{pmatrix}0\\1\end{pmatrix}$, in matrix notation. Second of all, it may seem a little confusing using a designation such as '0_z0_z' to mean '+1'. The sub indices z and z tells us that this value, +1, was obtained when ZZ was measured and found to be pointing along the positive z-axis. Conversely, for '0_z1_z' means '−1'. It tells us that this value −1 was obtained when ZZ was measured and found to be pointing along the negative z-axis. A similar line of reasoning applies to eigenvalues such as '0_z0_x'. Usually the numerical values of such states as '0_a0_b', '0_a1_b', '1_a0_b', or '1_a1_b' are ±1. Hence while the numerical value for any measurement can be the same, the way it was obtained can be very different and hence the meaning we associate with that value can change even though the numerical value doesn't.

[11] Consequently $|0_z0_x> = \begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}1\\1\end{pmatrix}/\sqrt{2}$, $|1_z0_x> = \begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}1\\1\end{pmatrix}/\sqrt{2}$. So we get for superpositions:

$(|0_z0_z> + |0_z1_z>)/\sqrt{2} = \left\{\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}1\\0\end{pmatrix} + \begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}0\\1\end{pmatrix}\right\}/\sqrt{2} = = \begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}1\\1\end{pmatrix}/\sqrt{2} = |0_z0_x>$ and for

$(|1_z0_z> + |1_z1_z>)/\sqrt{2} = \left\{\begin{pmatrix}0\\1\end{pmatrix}\otimes\begin{pmatrix}1\\0\end{pmatrix} + \begin{pmatrix}0\\1\end{pmatrix}\otimes\begin{pmatrix}0\\1\end{pmatrix}\right\}/\sqrt{2} = \begin{pmatrix}0\\1\end{pmatrix}\otimes\begin{pmatrix}1\\1\end{pmatrix}/\sqrt{2} = |1_z0_x>$.

[12] That is,

$(|0_z0_x> + |1_z0_x>)/\sqrt{2} = \left\{\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}1\\1\end{pmatrix} + \begin{pmatrix}0\\1\end{pmatrix}\otimes\begin{pmatrix}1\\1\end{pmatrix}\right\}/2 = \begin{pmatrix}1\\1\end{pmatrix}\otimes\begin{pmatrix}1\\1\end{pmatrix}/2 = |0_x0_x>$

$(|0_z0_x> + |1_z0_x>)/\sqrt{2} = |0_x0_x>$, eqn. 1I.

Using XX we find that

$XX\ |0_x0_x> = \text{'}0_x0_x\text{'}\ |0_x0_x>$. eqn. 1J.

MEMORY STATES OF QUANTUM AUTOMATA

It is important to remember that quantum automata do not hold in their memories eigenvalues of operators—they hold memories of what are called expectation values. Hence a quantum automaton holds in memory not the actual thing that was measured but a memory containing a meta-statement denoting a numerical value representing the thing that was measured.

The automaton's memory of each such state after a ZZ measurement will be written for eqns. 1A, 1B, 1C and 1D resp., are: $|\text{'}0_z0_z\text{'}>_{zz}$, $|\text{'}0_z1_z\text{'}>_{zz}$, $|\text{'}1_z0_z\text{'}>_{zz}$, and $|\text{'}1_z1_z\text{'}>_{zz}$. Here both quantum state vectors were measured along the z direction. We need to determine not just a value for a measurement but a meta-statement (in single quotes) telling us how that value was obtained. The absolute numerical value in each case is unity but clearly how that value was obtained is important as well, hence the sub-index sign, x or z, indicating the direction of the measurement.

For example suppose the quantum state vector for some system is $|\varphi>$. And suppose that a measurement of some observable, say **M**, is made when the system is in the eigenstate $|\varphi>$. The result will be $|\text{'}\varphi\text{'}>_m\ |\varphi>$.[13]

The result of that measurement will be dependent on just what observable is actually observed. But regardless of what observable is chosen we can ask the automaton to predict a value for *any* observable we wish, say the observable, **O**. We shall designate an operator P_o, which enquires about the state <u>predicted</u> and held in memory by the quantum automaton, i.e., P_o signifies the result obtained and predicted by the measurement of **O**. Of course to arrive at this prediction the automaton must be able to carry out whatever calculation is needed. The result will be in this general case,

$P_o\ |\text{'}\varphi\text{'}>_o = \text{'}<\varphi\ |\mathbf{O}|\ \varphi>\text{'}\ |\text{'}\varphi\text{'}>_o \equiv \text{'}O_{av}\text{'}|\text{'}\varphi\text{'}>_o$, eqn. 1K.

where '$<\varphi\ |\mathbf{O}|\ \varphi>$' is the average value of **O** (designated by O_{av}) when the state $|\varphi>$ is <u>not</u> an eigenvector of **O**. That is the best the quantum automaton can do under this circumstance. Of course if $|\varphi>$ is an eigenvector of some operator, say **F**, such that,

$\mathbf{F}|\varphi> = \text{'}f\text{'}\ |\varphi>$, eqn. 1L.

[13] This is standard von Neumann quantum physics boilerplate. See any textbook on quantum mechanics.

where 'f' is the eigenvalue for the quantum state vector $|\varphi>$, then we would find after a measurement of **F** has been carried out,

$$P_f\,|\varphi>= `<\varphi\,|\mathbf{F}|\,\varphi>'\,|`\varphi'>_f = `f'\,|`\varphi'>_f\,. \qquad \text{eqn 1M.}$$

In this case we have what we might call a "perfect" or "good" measurement. Consequently, considering the state $|`\varphi'>_f|\varphi>$, obtained after the **F** measurement we may (following Albert's seminal work) define an operator that determines how well the measurement was made. We shall call it an error measurement,

$$E_o \equiv P_o - \mathbf{O}. \qquad \text{eqn. 1N.}$$

Consequently we would find for **F** in this "perfect" case,[14]

$$E_f\,|\,`\varphi'>_f|\varphi> \equiv (P_f - \mathbf{F})\,|`\varphi'>_f|\varphi> = 0. \qquad \text{eqn. 1O.}$$

So whenever E_o is zero we would credit the automaton with having made a perfect measurement. The question remains, however, is the measurement an accurate one? For that we need to go a little deeper in the next section of the paper.

With this understanding we will see next, how quantum automata can do something that classical automata cannot do. They can contain self-referencing statements about themselves (or better about what they hold in memory, provided we use meta-language when dealing with measured values). My ideas here are largely based the pioneering work of David Z. Albert,[15] and later in this paper, on the work of Russian physicist A. V. Yurov[16].

A SHORT REVIEW OF ALBERT'S PAPERS SHOWING SELF-REFERENCE

In 1981, David Albert[17] proposed an ingenious scheme that enabled one to construct self-referring quantum states involving complex entanglements of an object and apparatus that he called a *quantum automaton* that measured the object. These complex states specifically involved entanglements of the automaton and an object with which it interacted. The automaton was viewed as having made a measurement on the object and, as well, self-measurements—those that involved the automaton performing a complementary measurement on itself during or after it had performed the given

---

[14] Here 'f' = f. Here the metavalue of f is f itself. That is the prediction matches perfectly with the value.

[15] Albert, David Z. *Op. cit.*

[16] Yurov, A. V. (Theoretical Physics Dept. Kaliningrad State University, Russia. yurov@freemail.ru.) "The Gödelzing Quantum-Mechanical Automata." arXiv: quant-ph/0301004v1. 3 Jan 2003.

[17] Albert, David Z. *op. cit.*

measurement on the object in question.

The scheme utilized or was based on the many-worlds interpretation[18] and perhaps for that reason alone, with the possible exception of quantum computer enthusiasts; little attention was paid to it. A key and brilliant insight[19] that Albert had came from realizing that if a quantum automaton could be constructed and if it operated along the lines of the many-worlds interpretation of quantum physics, then the memory of the apparatus could contain eigenvalues of both observables and complementary self-observing observables—in violation of the uncertainty principle—that is, it could contain information about the specific state of the object it measured and information about the state of itself while holding a superposition of the object's possible states entangled with the apparatus.

Because Albert's work was probably not widely known at the time and most likely because he didn't specify how such an automaton's memory could be constructed, future quantum computation theorists failed to follow up on his ground-breaking work. I believe I have discovered how such a memory could work (using the language of quantum computation[20]); it just may be that nature has already made quantum automata: the human brain operating as a self-referring quantum computer.

I repeat: the basic idea here is to make a difference between a purely algorithmic or logical process and what such a process can mean. Meaning occurs when a metamathematical statement is made about a mathematical statement. One may also substitute the world logical, algorithmic, physical, or any other measurable category. Hence meaning occurs when a metaphysical statement is made about a physical statement.

THE UNCERTAINTY PRINCIPLE EQUATIONS OF QUANTUM PHYSICS

First of all let us consider the simple two state system $|0_z0_z>$ just mentioned above in eqn. 1A and ask about operations involving ZZ and ZX. Consequently we have,

$$ZZ\,|0_z0_z> = |0_z0_z>, \qquad \text{eqn. 2A.}$$

and,

---

[18] See Bryce S. Dewitt. "Quantum mechanics and reality." *Physics Today* . Sept., 1970. p. 30-35. And see: Bryce S Dewitt and Neill Graham. *The Many-Worlds Interpretation of Quantum Mechanics*. Princeton, New Jersey: Princeton Univ. Press, 1973.

[19] One that David Deutsch acknowledged in his seminal paper on quantum computation. See: David Deutsch. "Quantum theory, the Church-Turing principle and the universal quantum computer." *Proceedings of the Royal Society of London*. Vol. A 400, pp. 97-117 (1985).

[20] Se for example, Nielsen, Michael A. and Isaac L. Chuang, *op cit.*

$$ZX\ |0_z0_z\rangle = |0_z1_z\rangle. \qquad \text{eqn 2B.}$$

A little calculation of the commutation relation shows that ZZ and ZX are incompatible,[21]

$$[\,ZZ; ZX\,]\ |0_z0_z\rangle \neq 0. \qquad \text{eqn. 2C.}$$

You cannot know both at the same time. This is essentially the uncertainty principle stated in terms of the commutation of two incompatible operators.

ACTION OF QUANTUM AUTOMATA

So given $|0_z0_z\rangle$, and if we measure ZZ, we get (where the super index (1) tells us that measurement has occurred, in this case along the zz direction),

$$|0_z0_z\rangle \rightarrow |0_z0_z^{(1)}\rangle \equiv |\text{'}0_z0_z\text{'}\rangle_{zz}\ |0_z0_z\rangle, \qquad \text{eqn. 3A.}$$

and if we had $|0_z1_z\rangle$, we would get,

$$|0_z1_z\rangle \rightarrow |0_z1_z^{(1)}\rangle \equiv |\text{'}0_z1_z\text{'}\rangle_{zz}\ |0_z1_z\rangle\ . \qquad \text{eqn. 3B.}$$

Suppose we have $|0_z0_x\rangle$ and we measure ZZ, we would get:

$$|0_z0_x^{(1)}\rangle \equiv (|0_z0_z^{(1)}\rangle + |0_z1_z^{(1)}\rangle)/\surd 2 = (|\text{'}0_z0_z\text{'}\rangle_{zz}\ |0_z0_z\rangle + |\text{'}0_z1_z\text{'}\rangle_{zz}\ |0_z1_z\rangle)/\surd 2. \qquad \text{eqn. 3C.}$$

We would then find:[22]

$$ZZ\ |0_z0_x^{(1)}\rangle = (\text{'}0_z0_z\text{'}|0_z0_z^{(1)}\rangle + \text{'}0_z1_z\text{'}|0_z1_z^{(1)}\rangle)/\surd 2. \qquad \text{eqn. 3D.}$$

To determine the accuracy of the automaton, we seek P_{zz} (prediction observable) and the error observable $E_{zz} \equiv (\,P_{zz} - ZZ\,)$. We find:

$$P_{zz}\ |0_z0_x^{(1)}\rangle = (\text{'}0_z0_z\text{'}\ |0_z0_z^{(1)}\rangle + \text{'}0_z1_z\text{'}\ |0_z1_z^{(1)}\rangle)/\surd 2, \qquad \text{eqn. 3E.}$$

and therefore:

$$E_{zz}\ |0_z0_x^{(1)}\rangle = 0. \qquad \text{eqn. 3F.}$$

Hence even though $|0_z0_x^{(1)}\rangle$ is an eigenstate vector of neither P_{zz} nor ZZ the automaton has made an accurate measurement of ZZ. The question is what does that mean in this case. It means that if we were to remeasure ZZ ourselves by performing a separate measurement on $|0_z0_x^{(1)}\rangle$ we would get the same result. However we would

---

[21] $ZX\otimes ZZ - ZZ\otimes ZX|0_z0_z\rangle = 2|0_z1_z\rangle \neq 0$.

[22] As well, $\text{'}0\ _z0_x\text{'} = +1$, etc. I emphasize not only the value implied by the operations upon their resp. quantum state vectors, but also how those values were obtain, as, e.g., a measurement along the z- and x-directions.

only realize that result in separate non-agreeing parallel universes or worlds.[23]

PREDICTIONS OF QUANTUM AUTOMATA

Suppose again we had $|0_z0_x\rangle$ and we measured ZZ. We got,

$$|0_z0_x^{(1)}\rangle = (|0_z0_z^{(1)}\rangle + |0_z1_z^{(1)}\rangle)/\sqrt{2}. \quad \text{eqn. 4A.}$$

We found:

$$E_{zz}|0_z0_x^{(1)}\rangle = 0. \quad \text{eqn. 4B.}$$

Now consider ZX $|0_z0_x^{(1)}\rangle$. Calculation of $P_{zx}\ |0_z0_x^{(1)}\rangle$ yields:[24]

$$P_{zx}\ |0_z0_x^{(1)}\rangle = 0. \quad \text{eqn. 4C.}$$

Consequently:

$$[P_{zz}; P_{zx}]\ |0_z0_x^{(1)}\rangle = 0, \quad \text{eqn. 4D.}$$

But since

$$[ZZ; ZX] \neq 0, \quad \text{eqn. 4E.}$$

we find that,

$$[E_{zz;}\ E_{zx}] \neq 0 \quad \text{eqn. 4F.}$$

This is garden variety quantum physics. We cannot know accurately both ZZ and ZX even though we can determine that $[P_{zz;}\ P_{zx}] = 0$. That is predictions of ZZ and ZX can't be made accurately for $|0_z0_x^{(1)}\rangle$ and regardless of the order of those predictions, the predicted results will be the same. But note since $P_{zz}\ |0_z0_x^{(1)}\rangle = ZZ|0_z0_x^{(1)}\rangle$ and $P_{zx}\ |0_z0_x^{(1)}\rangle = 0 \neq ZX\ |0_z0_x^{(1)}\rangle$, these predictions are not accurate. The automaton can make such predictions even though they both cannot be accurate.

SELF-REFERRING PREDICTIONS

Now suppose again we have the state $|0_z0_x^{(1)}\rangle$ and suppose we define an operator

---

[23] This all depends on what we mean by "re-measuring" the result we already obtained. In so doing we could actually change the memory states of the automaton.

[24] We get, ZX $|0_z0_x^{(1)}\rangle = (|`0_z0_z\text{'}\rangle_{zz}\ |0_z1_z\rangle + |`0_z1_z\text{'}\rangle_{zz}\ |0_z0_z\rangle)/\sqrt{2}$. Therefore $P_{zx}\ |0_z0_x^{(1)}\rangle = `\langle 0_z0_x^{(1)}|ZX|0_z0_x^{(1)}\rangle\text{'}\ |0_z0_x^{(1)}\rangle = \frac{1}{2}\ `({}_{zz}\langle `0_z0_z\text{'}|`0_z1_z\text{'}\rangle_{zz} + {}_{zz}\langle `0_z1_z\text{'}|`0_z0_z\text{'}\rangle_{zz})\text{'}\ |0_z0_x^{(1)}\rangle$. Taking each of these automata states as orthogonal, i.e., ${}_{zz}\langle `0_z0_z\text{'}|`0_z1_z\text{'}\rangle_{zz} = (10)\otimes(10)\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}0\\1\end{pmatrix}$ yields $P_{zx}\ |0_z0_x^{(1)}\rangle = 0$.

$\{ZX\}^{(1)}$ such that,[25]

$$\{ZX\}^{(1)}\ |0_z0_x^{(1)}> = `0_z0_x^{(1)}\text{'}\ |0_z0_x^{(1)}>, \qquad \text{eqn. 5A.}$$

Suppose we now measure $|0_z0_x^{(1)}>$ using $\{ZX\}^{(1)}$. We now get as a result:

$$|0_z0_x^{(2)}> = |`0_z0_x^{(1)}\text{'}>_{\{zx\}^{(1)}}\ |0_z0_x^{(1)}>, \qquad \text{eqn. 5B.}$$

and consequently,

$$P_{\{zx\}^{(1)}}\ |`0_z0_x^{(1)}\text{'}>_{\{zx\}^{(1)}} = `0_z0_x^{(1)}\text{'}\ |`0_z0_x^{(1)}\text{'}>_{\{zx\}^{(1)}} \qquad \text{eqn. 5C.}$$

Then we find: [26]

$$P_{\{zx\}^{(1)}}\ |0_z0_x^{(2)}> = \{ZX\}^{(1)}\ |0_z0_x^{(2)}> = `0_z0_x^{(1)}\text{'}\ |0_z0_x^{(2)}> \qquad \text{eqn. 5D.}$$

$$E_{\{zx\}^{(1)}}\ |0_z0_x^{(2)}> = (P_{\{zx\}^{(1)}} - \{ZX\}^{(1)})\ |0_z0_x^{(2)}> = E_{zz}\ |0_z0_x^{(2)}> = 0, \qquad \text{eqn. 5E.}$$

even though,

$$[ZZ; \{ZX\}^{(1)}]\ |0_z0_x^{(2)}> \neq 0, \qquad \text{eqn. 5F.}$$

$$[E_{\{zx\}^{(1)}}; E_{zz}]\ |0_z0_x^{(2)}> = 0. \qquad \text{eqn. 5G.}$$

In the state $|0_z0_x^{(2)}>$ we have the automaton holding a memory of two non-commuting observables ZZ and $\{ZX\}^{(1)}$. It can predict then both accurately. Both,

$$E_{zz}\ |0_z0_x^{(2)}> = 0, \qquad \text{eqn. 5H.}$$

and,

$$E_{\{zx\}^{(1)}}\ |0_z0_x^{(2)}> = 0. \qquad \text{Eqn. 5I.}$$

A 2nd automaton cannot do this because ZZ and $\{ZX\}^{(1)}$ are both, for it, observables of an external system, while the 1st automaton has eigenvalues for ZZ and $\{ZX\}^{(1)}$, both observables, contained within that system. Indeed if the 1st automaton

---

25 What would $\{ZX\}^{(1)}$ mean? One way to define it would be $(Z_zX_z)\otimes ZX$ where $(Z_zX_z)\ |`0_z0_z\text{'}>_{zz} = `0_z1_z\text{'}\ |`0_z1_z\text{'}>_{zz}$ and $(Z_zX_z)\ |`0_z1_z\text{'}>_{zz} = `0_z0_z\text{'}\ |`0_z0_z\text{'}>_{zz}$. Then $\{ZX\}^{(1)}\ |0_z0_x^{(1)}> = (Z_zX_z)\otimes ZX\ |0_z0_x^{(1)}> = `0_z0_x^{(1)}\text{'}\ |0_z0_x^{(1)}> = |0_z0_x^{(1)}>$.

26 This is a bit tricky. Since $ZX\ |0_z0_x> = \{\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\otimes\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\}\ \begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}1\\1\end{pmatrix}/\sqrt{2} = \begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}1\\1\end{pmatrix}/\sqrt{2}$ and therefore treating the automaton state as a quantum state vector, we find:

$$\left\{\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\otimes\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\right\}_{zx}\left\{\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}1\\1\end{pmatrix}/\sqrt{2}\right\}_{zx} = \left\{\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}1\\1\end{pmatrix}/\sqrt{2}\right\}_{zx}.$$

So we get eqn. 5C where the value of $`0_z0_x^{(1)}\text{'} = 1$.

would share its information with a 2nd automaton, it would change that information from certainty to a probability.[27]

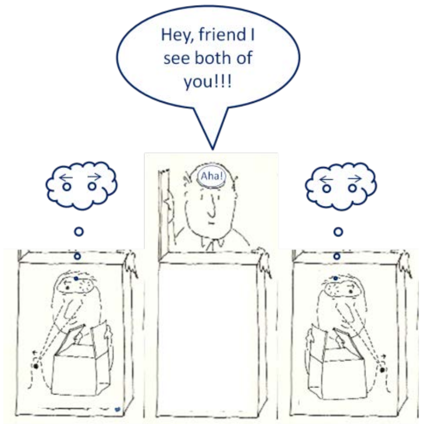

Figure 3. Professor Wigner holds both possible friend's observations.

SECRET MEMORIES IN VIOLATION OF THE UNCERTAINTY PRINCIPLE

It is also possible to consider other memory states, even for this simple system. Consider $|0_z0_x^{(2)}>$ and $|1_z0_x^{(2)}>$. We can consider the superposition of these two quantum state vectors and, in eqn. 6C, an appropriate operator, $\{XX\}^{(2)}$ where,

$$|0_x0_x^{(2)}> = (\ |0_z0_x^{(2)}> + |1_z0_x^{(2)}>\ )/\sqrt{2}, \qquad \text{eqn. 6A.}$$

and where,

$$|1_z0_x^{(2)}> = |`1_z0_x^{(1)}{}'>_{\{zx\}^{(1)}}\ |1_z0_x^{(1)}>. \qquad \text{eqn. 6B.}$$

We can now consider measuring $|0_x0_x^{(2)}>$ using $\{XX\}^{(2)}$ where,

$$\{XX\}^{(2)}\ |0_x0_x^{(2)}> = `0_x0_x^{(2)}{}'\ |0_x0_x^{(2)}> \qquad \text{eqn. 6C.}$$

Thus $\{XX\}^{(2)}$ plays the same role for $|0_x0_x^{(2)}>$ that $\{ZX\}^{(1)}$ plays for $|0_z0_x^{(1)}>$ and

[27] To see how this works consider a 2nd automaton asking the 1st automaton about what it has measured for both $\{ZX\}^{(1)}$ and ZZ concerning the state $|0_z0_x^{(2)}>$. In order to ask the 2nd automaton, it must also interact with the system containing both the 1st automaton and the outside system it had interacted with. In effect it would need to carry out measurements of $\{ZX\}^{(1)}$ and then ZZ in that order or in the reverse order. Labeling the 2nd automaton with a superscript 2 and the 1st with a superscript 1, we have, ${}^{2}|`0_z0_x^{(1)}{}'>_{\{zx\}^{(1)}}\ {}^{1}|`0_z0_x^{(1)}{}'>_{\{zx\}^{(1)}}\ [{}^{2}|`0_z0_z{}'>_{zz}\ {}^{1}|`0_z0_z{}'>_{zz}\ |0_z0_z> + {}^{2}|`0_z1_z{}'>_{zz}\ {}^{1}|`0_z1_z{}'>_{zz}\ |0_z1_z>\ ]/\sqrt{2}$ $\neq |0_z0_x^{(2)}>$. If we had made the same request to the 1st automaton it would yield the same quantum state vector $|0_z0_x^{(2)}>$ since it already had made these measurements.

$|1_z0_x^{(1)}>$.[28] We can follow the same measurement procedure to arrive at:

$$|0_x0_x^{(3)}> = |`0_x0_x^{(2)}\text{'}>_{\{XX\}^{(2)}}\ |0_x0_x^{(2)}> \qquad \text{eqn. 6D.}$$

In principle we could carry this as far as we wish depending on the storage capacity of the automaton. If we consider what the quantum state vector $|0_x0_x^{(3)}>$ tells us the story is fascinating. We can imagine the automaton containing memory cells or if you wish bundles of neurons. In one cell we have the memory, '$0_x0_x^{(3)}$'. In a 2^{nd} cell we have the memory '$0_x0_x^{(2)}$', in a 3^{rd} cell, '$0_z0_x^{(1)}$', and in a 4^{th} cell, in one Everett world, '0_z0_z', and in the other parallel universe, '0_z1_z'. Even though none of these self-referring observables (i.e., ZZ, $\{ZX\}^{(1)}$, and $\{XX\}^{(2)}$) commute with each other, the automaton "knows" this "classical" information about what it has measured in the outside world (i.e., ZZ) and what it has measured about itself (i.e., its own self-referring memories of $\{ZX\}^{(1)}$, and $\{XX\}^{(2)}$).

Indeed if we had chosen to begin with sixteen base states such as $|0_z0_z0_z0_z>$ instead of $|0_z0_z>$ we would find eigenvalues for $|0_z0_z0_z0_z>$, $|0_z0_z0_z0_x^{(1)}>$, $|0_z0_z0_x0_x^{(2)}>$, $|0_z0_x0_x0_x^{(3)}>$, $|0_x0_x0_x0_x^{(4)}>$, and so forth. Hence the automaton's memory could hold eigenvalues for the observables ZZZZ, $\{ZZZX\}^{(1)}$, $\{ZZXX\}^{(2)}$, $\{ZXXX\}^{(3)}$, and $\{XXXX\}^{(4)}$, simultaneously with $E_{zzzz} = E_{\{zzzx\}^{(1)}} = E_{\{zzxx\}^{(2)}} = E_{\{zxxx\}^{(3)}} = E_{\{xxxx\}^{(4)}} = 0$, when operating on $|0_x0_x0_x0_x^{(4)}>$, all in violation of the uncertainty principle.

1ST CONCLUSION

What does this all mean? It means the quantum automaton holds information about its own memories as well as about its external observations. This kind of self-reference means the automaton may be exhibiting a model of our own subjective self-referencing experience: namely an exhibition of the privacy of consciousness of memory. In principle assuming human memory works this way, each of us holds private conceptions not only of our observations made in the outside world, but also "secret" conceptions of how we think about those observations—that is observations of our inside or subjective "meta" world. Surprisingly we each cannot hold these private self-referencing observations if we choose to share them with others. In so doing we actually change the values of our memories. Perhaps this explains how we can learn new values—we simply tell others what old values we hold. It may also explain why it helps to "talk thing out" with a good friend or psychologist; in so-doing we change our "secret" memories.

---

[28] What would $\{XX\}^{(2)}$ be? A similar line of reason that gave $\{ZX^{(1)}\} = (ZX)_{zz} \otimes ZX$ gives us $\{XX\}^{(2)} = \{(XX)_{zz} \otimes XX\}_{\{ZX\}^{(1)}} \otimes \{XX\}_{zz} \otimes XX$. As can be seen after a little algebra $\{XX\}^{(2)}\ |0_x0_x^{(2)}> = `0_x0_x^{(2)}\text{'}|0_x0_x^{(2)}> = |0_x0_x^{(2)}>$.

Of course we may choose not to reveal ourselves to others. Thus "secret" information we each hold may indeed play a large role on how we think about ourselves and others and why it is so difficult for each of us to "walk in the other's shoes" before making judgments.

PART II: RUSSIAN GÖDELIZATION

Because of the automaton's ability to hold self-referencing information, a natural question arises with its connection with the famous consistency-completeness proofs of Kurt Gödel in which Gödel showed that within any formal and logically consistent system statements (proofs) can be formulated that cannot be proven within the system. These statements invariably are found to be self-referring statements. Consequently Gödel's proof may have ramifications for the ability of quantum automata to hold self-referring information about their own memories in violation of the uncertainty principle (i.e., outside the formal system of quantum mechanics). Gödel's proof may help us understand why such a "self-conscious" automaton can "know" statements within the formal system of quantum mechanics that are not provable within that objective system. Penrose calls this ability to see the truth of such a self-referring statement, **Gödelization.**[29] It appears apparent that we humans can Gödelize as easily seen be inspecting Figs. 2 and 4. Your understanding of these "unprovable situations" means you (I presume you are a human being) have performed a Gödelization.

Does human consciousness depend on our mental abilities to hold such "non-provable" information in memory? Is Gödelization necessary and sufficient for human consciousness? Next we will explore these questions.

Perhaps the first to consider these questions in the light of quantum physics was Russian physicist, Artyom Yurov. In his interesting paper, 30 Yurov asks:

- Can minds can do what automatons cannot do?
- Can minds transcend formalized rules?
- Do we need new physics to understand "the mind"?

Yurov then goes on to consider these question by putting Gödel's proof in the language of operations on quantum state vectors.

[29] Penrose, Roger, *op.cit.*

[30] See Yurov, A. V. Theoretical Physic Dept. Kaliningrad State University, Russia. yurov@freemail.ru "The Gödelzing Quantum-Mechanical Automata." arXiv: quant-ph/0301004v1, 3 Jan 2003.

GÖDEL IN BRIEF

Let me start by briefly explaining the relevancy of Gödel's proof to quantum physics. Erik van Heusden, in his paper dealing with undecidibility in physics,[31] presented us with some Gödelization issues. In essence Gödel showed:

- Any consistent formal system, say F, within which a certain amount of elementary arithmetic can be carried out is <u>incomplete</u>; i.e., there are statements in the language of F which can neither be proved nor disproved in F.
- For any <u>consistent</u> system, F, within which a certain amount of elementary arithmetic can be carried out, the consistency of F cannot be proved in F itself.

What about Nature itself? If nature may be represented as some formal system, then there are four options that can be roughly categorized as follows:

(1) The fundamental laws of quantum physics cannot be represented as a consistent formal system.

(2) The fundamental laws of quantum physics are not subject to the Gödel's incompleteness theorems.

(3) Any axiomatic representation of the laws of quantum physics requires an infinite number of axioms.

(4) Real phenomena exist that do not follow from the laws of quantum physics.

Moreover, the consistency of the laws of quantum physics is one of these phenomena.

Heusden asks us to make a choice after considering several arguments for and against. I do not ask you to make such a choice. I am somewhat torn between option (4) and option (1). I have attempted to show here dealing with self-referencing quantum automata that a meta-language representation of eigenvalues of appropriate quantum state vectors can be viewed as part of quantum physics–albeit a part that before Albert's papers was not suspected. One may conclude that I have gone somewhat beyond Albert's views here in bringing in metavalues—signs that point to a meaning of the value obtained in a measurement. Perhaps one may argue that I have put semiotics into the fold.

Before we get into that consider how John D. Barrow analyzed Gödel's proofs in

---

[31] van Heusden, Erik F. G. "On Undecidibility and the Laws of Physics" *Journal of Physics: Conference Series* 701 (2016) 012025.

the light of physics.[32] He pointed out in an easily understood example that the careful study of axiomatic systems revealed that even Euclid's beautiful development of plane geometry made use of unstated axioms. In 1882, Moritz Pasch gave a very simple example of an intuitively "obvious" property of points and lines that could not be proved from Euclid's classical axioms. If the points A, B, C, and D lie on a straight line such that B lies between A and C and C lies between B and D then it is not possible to prove that B lies between A and D. The picture of the set-up (see below) made it appear inevitable but that is not a substitute for a proof.

Figure 4. Euclid's unprovable line.

Hence here we have an example of an obvious property of points on a straight line that cannot be proven within the closed and consistent system of Euclid's geometry. I repeat, Euclid's geometry is quite consistent and closed. More precisely, a system is *consistent* if we cannot prove that a statement S and its negation, ~S, are both true theorems. It is *complete* if for every statement S we can form in its language, either S or its negation, ~S, is a true theorem. It is *decidable* if, for every statement S that can be formed in its language; we can prove whether S is true or false. Thus, if a system is decidable it must be complete.

Gödel proved that any system rich enough to contain arithmetic must be incomplete and undecidable! Briefly, here's Gödel's symbolic proof:

First, Gödel showed that each mathematical formula, like '0 = 0', can be given a unique number, the Gödel number. Gödel used a rather complex way to do this, but he began with some simple numbers. For example the Gödel number for '0' is 6 with other low numbers for signs like '∃' (which means there exists) is 4, '~' (the negation sign meaning not) is 1, the punctuation marks: left parenthesis, '(' is 8, right parenthesis, ')' is 9, period, '.' is 10 and, equal sign, '=' is 5. It is also possible to assign Gödel numbers to statements such as the equation, '0=0'. To do so Gödel took the prime numbers, as many as he needed such as 2, 3, 5, etc., and raised each prime to a power equal to the Gödel number of the sign in the equation. Here those numbers are 6, 5, and 6 resp. So that the equation, '0=0', has the unique Gödel number,

[32] Barrow, John D., "Gödel and Physics." In Matthias Baaz (ed.), *Kurt Gödel and the Foundations of Mathematics: Horizons of Truth*. Cambridge University Press. pp. 255 (2011).

$$2^6 \times 3^5 \times 5^6 = 64 \times 243 \times 15{,}625 = 243{,}000{,}000. \qquad \text{eqn. 6E.}$$

No other mathematical statement has this Gödel number. Once one has the Gödel number for any arithmetical statement it is possible to recover or decode it to realize the original statement.

Second, Gödel considered numbering statements about mathematical statements—so-called metamathematical statements as I discussed above. By doing so it is possible to express statements like 'the sequence of formulas A are a proof of formula B' as an arithmetical relation between the Gödel numbers for A and B. Thus meta-mathematics can be mapped into arithmetical statements or proofs. The idea is since every mathematical statement has a unique Gödel number, then any metamathematical statement about a sequence of statements constituting a relation between them can be construed to be a relationship between their Gödel numbers. Thus find the relation between the Gödel numbers and you prove the relation between the statements.

Thirdly and lastly, included in the list of possible metamathematical statements is one we will call G which is the statement about G itself namely: "The statement G cannot be derived or proven from the axioms of mathematics." Of course we may have many G statements with Gödel numbers, G_i that states: "The statement G_i cannot be derived or proven from the axioms of mathematics" where i can run to infinity.

So if we consider a self-referring Gödel statement, G, what shall we conclude about its consistency? Suppose that G could be demonstrated within the mathematical axioms. Then the axioms must be inconsistent because one could both demonstrate G and show that it cannot be demonstrated. Hence any mathematical system that proves a false statement (such as $0=5$) and its true negation ($0\neq5$) must be inconsistent. On the other hand, if G can't be demonstrated, then G is true. By the mapping of meta-statements into Gödel numbers, G corresponds to a true relation between these Gödel numbers, but one which cannot be deduced from the axioms. Thus mathematics is either inconsistent or incomplete. Since mathematics seems to be consistent, so far, one may surmise that mathematics is incomplete.

YUROV'S GÖDELZING QUANTUM AUTOMATA

Yurov used the language of Gödel's proof to show how self-referential statements that are not provable within that objective system of quantum physics, such as the "collapse of the quantum wave function," can be proven using Albert's quantum automata. Gödelization as I discussed above is a process that enables us to understand proofs of statements that lie outside of the formal system that contain them. Since quantum physics is a formal mathematical system, therefore, it should be subject to Gödel's proofs of incompleteness.[33] Hence a proposition proving what occurs when a measurement occurs and mind records it and understands its meaning, must lie outside that system. Yurov ask us to ponder if quantum automata can Gödelize information—something that Penrose suggests can only be done by minds outside of quantum physics.

Yurov has us consider any propositional function, $s_k(w)$, where k is the proposition's Gödel number, applied to a number, w (with the sign 'w' assumed to have Gödel number 17). For example, the proposition, 'x=w', has a quite large Gödel number, $k = 2^{11} \times 3^5 \times 5^{17} = 3{,}796{,}875 \times 10^{11}$. Suppose a string of propositions, U(n) [that constitute some proof of a proposition such as $s_k(w)$], has the Gödel number, n. Accordingly one should be able to find a proof of $s_k(w)$ within a classical (arithmetical) system, W_c, using U(n). So that the statement,

$$s_k(w) = \exists\, n\ [U(n) \text{ proves } s_k(w)], \qquad \text{Eqn. 7A1.}$$

which says it is true that there exists a particular Gödel number n, such that U(n) constitutes a proof of $s_k(w)$.

But according to Gödel, within a classical (arithmetical) system, W_c, there does not exist a proof such that for any number n, U(n) proves $s_w(w)$.

In symbolic logic language:

$$s_w(w) = \sim \exists\, n\ [U(n) \text{ proves } s_w(w)] \qquad \text{eqn. 7A2.}$$

That is the proposition, $s_w(w)$, with Gödel number, w, cannot be demonstrated by any string of propositions U(n) within W_c, which is based on arithmetic propositions. Hence any proposition with the Gödel number, w, about the number, w, cannot be proven within the system, W_c.

Now suppose we have a quantum automaton that proves theorems in W_c. It does so by putting (translating) classical or arithmetical propositions into the language of quantum physics. This is no more mysterious than having a quantum computer that solves numerical problems.

[33] We may guess here that the 1935 incompleteness paper by Einstein and his colleagues at Princeton may have been influenced by Gödel who visited Einstein many times while at Princeton. See Einstein, Albert; Podolsky, Boris; and Rosen, Nathan. "Can The Quantum-Mechanical Description of Physical Reality Be Considered Complete?" *Physical Review*. Vol. 47 (1935), p. 777.

Let |w; k> denote a quantum state vector that corresponds to the mathematical statement $s_k(w)$ within W_c. Again we may think about 'x=w' as an example.

Correspondingly, for |w; k> we can find an operator-observable,[34] $S_k(w)$, such that,

$$S_k(w)\ |w; k\rangle = \text{'k'}\ |w; k\rangle. \qquad \text{eqn. 7B1.}$$

This means that when $S_k(w)$, a function of the number w, and Gödel number, k, operates on the quantum state eigenvector |w; k>, the result is the eigenvalue, 'k', which you will remember is the Gödel number of $s_k(w)$. So if we can determine k we can determine $s_k(w)$.

But to do so means we need to go beyond W_c to a a new system W_q (meaning quantum system), which we take to have a list of quantum gates U_j, which can be sequenced. This leads to or constitutes a proof of $s_k(w)$. That is:

$$U(w; k) = U_1 \otimes U_2 \otimes U_3 \otimes U_4 \otimes \ . \ . \ . \qquad \text{eqn. 7B2.}$$

So given the sequence of gates, U(w; k), we can always find a proof of $s_k(w)$ (about the number w) by following a simple procedure to determine |w; k> from any initial state |φ> provided we construct U accordingly, so that,

$$U(w; k)\ |\varphi\rangle = |w; k\rangle. \qquad \text{eqn. 7C1.}$$

In essence, by passing |φ> through an appropriate sequence of gates as in eqn. 7B2, we arrive at the state, |w; k>. Whereupon we can always find the Gödel number k from calculating,

$$\langle\varphi|\ U^\dagger S_k(w)\ U\ |\varphi\rangle = \text{'k'}, \qquad \text{eqn. 7C2.}$$

and then given k we can determine (prove) the proposition $s_k(w)$.

This is realizable if,

$$U(w; k)\ |w; k\rangle = \text{'u(w; k)'}\ |w; k\rangle, \qquad \text{eqn. 7C3.}$$

(where the value of 'u(w; k)' is 1) for then the commutator,

$$[S_k(w), U(w; k)] = 0. \qquad \text{eqn. 7C4.}$$

That is, within the system W_q both $S_k(w)$ and U are capable of being known simultaneously. In plain language within the system, W_q a proof of $s_k(w)$ can be given by U(w; k) so long as $k \neq w$.

What about $s_w(w)$? Since W_q and W_c are both arithmetic systems we shouldn't expect that we would be able to prove $s_w(w)$ within W_q and W_c. Remember $s_w(w)$ has a

---

[34] Given any quantum state vector $|v_i\rangle$ it is always possible to find an operator V for which $V\ |v_i\rangle = v_i\ |v_i\rangle$.

Gödel number w, and states that there cannot be a proof of any statement $s_w(w)$ within W_c or W_q. Correspondingly, it should be the case that,

$$[S_w(w), U(w; k)] \neq 0. \qquad \text{eqn. 7C5.}$$

Else $s_w(w)$ could be proven within W_q in violation of Gödel's proof. So let's take a look.

Following Yurov ad Albert, let's only consider statements about the Gödel proposition $s_w(w)$ characterized within W_q by self-referring observables $S_w(w)$ which we write simply as s_w and S_w, resp. Here we assume S_w is an observable that implies or states the Gödel proposition, s_w. Can it do so within W_q?

To keep things as simple as I can, suppose we have just two self-referring "Gödel" quantum state vectors, $|w_1>$ and $|w_2>$.[35] and suppose we have the state $|Y>$ where (as we showed in eqn. 1E),

$$|Y> = (|w_1> + |w_2>)/\sqrt{2} \qquad \text{eqn. 7D1.}$$

And suppose, writing U instead of U(w; w),

$$U\ |Y> = \text{`Y'}\ |Y>. \qquad \text{eqn. 7D2.}$$

Then,

$$S_w\ |w_1> = \text{`}w_1\text{'}\ |w_1>, \qquad \text{eqn.7D3.}$$

and

$$S_w\ |w_2> = \text{`}w_2\text{'}\ |w_2>. \qquad \text{eqn. 7D4.}$$

Also we find,

$$S_w\ |Y> = (\text{`}w_1\text{'}\ |w_1> + \text{`}w_2\text{'}\ |w_2>)/\sqrt{2}. \qquad \text{eqn.7D5.}$$

Now U and S_w are incompatible. We cannot know both so we should find,

$$[S_w; U] \neq 0. \qquad \text{eqn. 7D6.}$$

Now let's bring in measurement as we did before. When a measurement of S_w is completed on $|w_1>$, we have,

$$|w_1^{(1)}> = |\text{`}w_1\text{'}>_{sw}\ |w_1>, \qquad \text{eqn. 7E1.}$$

and similarly for $|w_2>$,

$$|w_2^{(1)}> = |\text{`}w_2\text{'}>_{sw}\ |w_2>. \qquad \text{eqn. 7E2.}$$

We also can have prediction variables such as P_{sw} such that,

$$P_{sw}\ |\text{`}w_1\text{'}>_{sw} = \text{`}w_1\text{'}\ |\text{`}w_1\text{'}>_{sw}, \qquad \text{eqn. 7E3.}$$

and,

[35] For example, $|w_1>$ would point to the self-referring statement $s_{w1}(w_1)$ that has Gödel number w_1.

$$P_{sw}\ |\text{'}w_2\text{'}>_{sw} = \text{'}w_2\text{'}\ |\text{'}w_2\text{'}>_{sw}.$$ eqn. 7E4.

Hence using E_{sw} as we did in eqn. 1N,

$$E_{sw}\ |w_1^{(1)}> = (P_{sw} - S_w)\ |w_1^{(1)}> = 0,$$ eqn. 7E5.

and similarly,

$$E_{sw}|w_2^{(1)}> = 0.$$ eqn. 7E6.

So if we have the state,

$$|Y^{(1)}> = (|w_1^{(1)}> + |w_2^{(1)}>)/\surd 2,$$ eqn. 7E7.

then also,

$$E_{sw}|Y^{(1)}> = 0.$$ Eqn. 7E8.

Now reconsider the state $|Y^{(1)}>$. Even though all commutators with P_{sw} and P_u vanish,

$$[P_{Sw}; U] = [P_{sw}; S_w] = [P_{sw}; P_u] = [P_u; S_w] = [P_u; U] = 0,$$ eqns. 7E9.

when operating on $|Y^{(1)}>$, we see that nevertheless,

$$[E_u; E_{sw}]\ |Y^{(1)}> \neq 0,$$ eqn. 7F1.

because,

$$[S_w; U] \neq 0,$$ eqn. 7F2.

which is in accord with Gödel's proof and good quantum physics. Hence simultaneous predictions of U and S_w are not accurate. But now consider the self-referring state,

$$|Y^{(2)}> = |\text{'}Y^{(1)}\text{'}>_{sw}\ |Y^{(1)}>,$$ eqn. 7F3.

and suppose we have an operator $U^{(1)}$ such that,

$$U^{(1)}\ |Y^{(1)}> = \text{'}Y^{(1)}\text{'}\ |Y^{(1)}>.$$ eqn.7F4.

Since we can see that,

$$S_w\ |Y^{(2)}> = P_{sw}\ |Y^{(2)}> = |\text{'}Y^{(1)}\text{'}>_{sw}\ (\text{'}w_1\text{'}\ |w_1^{(1)}> + \text{'}w_2\text{'}\ |w_2^{(1)}>)/\surd 2.$$ eqn. 7F5.

We find,

$$E_{sw}\ |Y^{(2)}> = 0,$$ eqn. 7F6.

and then we also see that,

$$P_{u^{(1)}}\ |Y^{(2)}> = \text{'}Y^{(1)}\text{'}\ |Y^{(2)}>,$$ eqn. 7F7.

so,

$$E_{u^{(1)}}\ |Y^{(2)}> = 0.$$ eqn. 7F8.

Since the automaton can "know" both $U^{(1)}$ and S_w we have,

$$[E_{u}^{(1)}; E_{sw}]\ |Y^{(2)}> = 0. \qquad \text{eqn. 7F9.}$$

Even though,

$$[U^{(1)}; S_w]\ |Y^{(2)}> \neq 0, \qquad \text{eqn. 7G1.}$$

these observables cannot be exactly known externally to the automaton; it can hold this information internally—it can Gödelize.

Remember that S_w refers to a self-referring statement, s_w, while $U^{(1)}$ refers to a proof of s_w labeled by the Gödel number '$Y^{(1)}$'.

Does this mean the automaton can prove s_w within an augmented system, $W_q^{(1)}$ that takes into account self-referring states? No, it cannot do so because to do so it must externalize (couple with another external automaton, e.g.) the information it holds and that would necessarily change that information. It just means the quantum automaton "knows" both Gödel numbers '$Y^{(1)}$' and 'w_1' in one world and '$Y^{(1)}$' and 'w_2' in the other.

So do we need new physics to Gödelize? From one point of view we don't: The automaton holds both the system $W_q^{(1)}$ represented by '$Y^{(1)}$' and the self-referring propositions represented by 'w_1' in one world and 'w_2' in the other. So it seems that we have Gödelized the information.

From a 2nd point of view, we do: If Gödelization is an unalgorithmic procedure (as I suggest here by using meta-statements) then it also appears that we have a post-quantum physics with $W_q^{(1)}$ that contains an unsuspected possibly unalgorithmic procedure.

The question still remains that even though quantum automata can contain information in violation of the uncertainty principle (hence constitute a kind of post-quantum physics) they cannot tell anyone what they "know" without changing the information they hold. So doing would put their information in compliance with the uncertainty principle.

2ND CONCLUSION: SELF-REFERRING CONSCIOUSNESS

Here things get quite interesting from the point of view of describing a quantitative difference between *subjective* and *objective* points of view. The automaton is perfectly capable of "knowing," that is, simultaneously holding in memory eigenvalues of non-commuting observables in which, as I have said, such knowledge is in violation of the uncertainty principle. However, these are rather strange non-commuting observables—not the kind you usually find in grand-dad's quantum mechanics—when you take into account that they not only refer to a state corresponding to an outside object, but also to the self-referring measurement of the state of the recording device containing that information. Hence we have a rather unique situation here. As Albert pointed out; it seems that the state of knowledge of the automaton depends on its identity! Thus we can postulate the speculative axiom that:

> *Axiom: It is necessary and sufficient that a violation of the uncertainty principle occurs within the boundaries of, and in order for, a system including its memory to possess self-identity.*

Or we can put it in another way:

> *Alternative Axiom: In order for a system and its memory to possess self-identity, that is, exhibit self-consciousness, a violation of the uncertainty principle must occur within the given system's memory.*

What this tells us is rather amazing, if you think about it. Assume for the moment, that by the measurement-recording abilities of the automata described herein, I mean the actions taken by the mind of an observer or perhaps an AI device in recording a memory of an event. What does the above axiom indicate about the way our minds take in information? By violating the uncertainty principle it appears that we have returned to a classical world description—one where the actions of mind seemingly play no role in their effect upon the objects being observed and measured. However, one shouldn't think that such a violation indicates a return to classical physics,[36] but, instead, that in attempting to obtain a description of the world wherein we seem to be separated from it, our minds must operate, perceive, or hold memory *as if* classical physics pertained. That is, our minds perceive the world without perceiving themselves to be part of it.

In the world of the self-referring mind, where knowledge of eigenvalues of non-commuting observables can apparently be accessed (provided that knowledge includes self-knowledge), objects (including the self) appear to have well-defined values just as the objects of a classical world do appear to us. But there is more to this than just the appearance of classical values.

The *identity* of a system (object plus automaton) depends on *it* holding this information intact. As long as it does so, it maintains and possesses a unique identity—that is, a distinction between it and another—for it holds information that cannot be accessed by an outside automaton/observer without disruption, in an unpredictable manner (although one could make a probability prediction), of the information it

[36] Hardly at all! For in classical physics, we have the complete absence of anything like an observer.

contains. Moreover, attempts to dislodge such information would continually alter, as it were, the system's "information" boundary—the very distinction a system would make between itself and its objective environment.

Hence a distinction between the system, and the objective world it may relate to, arises chiefly from its ability to hold on to this undisclosed information. It holds this information, so to speak, in secret and in its secret holding it becomes aware of itself as distinct from any others that do not have privy to its holding. Its ability to be a separate *thing* implies the holding of "secret" knowledge. Hence, if I make a further leap here, objectivity (the mindful distinction of separate objects) arises from subjectivity—a holding of secret knowledge.

Also, please note, the knowledge that we normally take to be held within the confines of the automaton's memory cannot simply be said to just reside there. It must reside in the object-automation system, for this memory involves both the automaton's multiply-reflected, uncertainty principle-violating, memory and the state of the thing observed in the peculiar manner that is indicated by any of its states.

In this rather simple manner, I believe I have explained how it is that our experiences of the "out there" world, that we know must include our memories in order to be perceived, appear to us as occurring "out there" even though we strongly suspect from neurophysiology that such experiences must be projected from our brains and nervous systems in some yet to be determined manner. Hence, assuming my memory works this way, I may know more about an object I perceive, that is, I may have recorded more about this object, than I can possibly disclose, even to myself! Hence the world may appear classical to me provided I don't (or cannot) disclose all that I have perceived about it in the past.

I believe that in this model we see both how it is that we perceive a classical world, or have the predilection to imagine and to perceive a classical order in that world, and how that world appears to be outside of us in spite of neurophysiological evidence indicating that we "should" be perceiving our perceiving instruments and not the results of those instruments' perceptions.[37]

Of course any attempt by an automaton to export any of this self-held knowledge

[37] Of course this statement is debatable. For just what should a person perceive or judge a perception to be? his question brings up the nested Chinese box version of consciousness wherein whatever box we put an observer, he will always be looking up and out at the box preceding the box that holds himself. We don't have such an experience. As we peer through our eyes we have no indication that we are located inside of our heads. For if we did, we would then be perceiving the backs of our retinas, or the drums of our ears, or some other sense organ in the same way that we look at a voltmeter to tell us about the voltage state of an electrical circuit.

to an outside observer or 2nd automaton alters this knowledge and renders any further information subject to uncertainty while simultaneously making the information no longer accurately remembered. Perhaps the implantation of false memories proceeds by such a mechanism.

Here we are in the paradoxical situation of recognizing that a quantum automaton knows more than it can ever tell. If we attempt to ask it what it “knows,” that is, attempt to pull the complementary values or information it holds from it by measuring its simultaneously held complementary eigenvalues or coupling it to an outside automaton, it will in effect render the information it gives inaccurate or no longer valid. It can say what state it was in, but once the information has been requested by another automaton, it cannot accurately say what eigenstate it currently holds in memory or even what state the system is in without losing the information.

Again this is rather startling if you think about it as well. For in this simple manner, I believe I also have explained how it is that each of us holds in memory a sense of privacy of consciousness; we each see a separate and objective world and cannot disclose that private view without disruption. In this way we see how natural it is that mind appears to be divided into separated systems, that is, minds.

Perhaps this model may help us understand human memory and consciousness, as well as providing a way for quantum automata to operate within AI quantum computing devices involving and not involving self-consciousness, both in a sentient being and in an AI device.

Founder and Director of *Have Brains/Will Travel*: a Global Quantum Physics Educational Company, San Francisco CA, USA
fred@fredalanwolf.com