Computation, Consciousness and the Quantum

Bruno Marchal, IRIDIA, Brussels University

September 24, 2000

Abstract: It is sometimes said that Everett's formulation of Quantum Mechanics dispenses us with the need of a theory of consciousness in the foundation of physics. This is false as Everett himself clearly recognized in his paper. Indeed he has build its quantum mechanics formulation by using explicitly the mechanist or computationalist hypothesis in psychology. Everett and his followers have then derived the subjective appearance, in the mind of machine-observers, of indeterminacy and non-locality from the Schrödinger Equation. I argue in this paper that if we take the computationalist hypothesis seriously enough then the Schrödinger equation *itself* should be derivable from the computationalist theory of consciousness, making ultimately physics a branch of machine's psychology. I sketch the basic argument and illustrate it with two embryonic derivations. In some sense I criticize Everett for his lack of radicality.

1 Quantum Realism

Let me put it in this way: *all* sufficiently realist[1] interpretations of quantum mechanics accept the existence of parallel *situations*. This follows from the superposition principle which comes itself from the experimental evidence with (single) particle(s) interference phenomena. The parallelism is maintained for some time, at least for the isolated system, thanks to the linearity of the Schrödinger equation, i.e. the linear evolution of the system. With this in mind we can classify the realist interpretations of Quantum Mechanics by the degree of contagion of the parallelism/superposition.

If we accept, like Everett, the universal validity of Schrödinger equation, then the contagion is universal [12, 13]. Anything which interacts properly with a superposition will be in a superposed state :

[1] I omit in this paper the positivist attitude which considers that science provide only rules for predicting numbers (the *don't ask* interpretation).

$$|a_i\rangle(\alpha|b\rangle + \beta|c\rangle) = \alpha|a_b\rangle|b\rangle + \beta|a_c\rangle|c\rangle$$

where $|\alpha|^2 + |\beta|^2 = 1$, and $|a_i\rangle|d\rangle \rightarrow |a_d\rangle|d\rangle$ describes the interaction. This is a simple consequence of the linearity of the tensorial product.

Accepting, with Everett, the universal validity of the Schrödinger equation makes Quantum Mechanics an entirely deterministic, local, realist, complete and *democratic* theory. By "democratic" I mean that the observer has no particular status and can be described by the laws he is inferring from his observation of nature : no *Heisenberg cut*. We loose something though. We loose the principle of *absolute* unicity of outcomes in measurement. But then the theory is able to explain the *apparent* unicity of outcome in the memory of the multiplied observers, and, with the work on decoherence (as Everett has begin to explain) we even get an explanation of the emergence of the classical mechanics *experience* as shared by the average observers.

Nevertheless some people still insist on keeping the principle of *absolute* unicity of outcomes in measurement. They must hope that the linear contagion of the parallelism will stop somewhere in between themselves and the coherent state they are observing. If that happens, the *very* observation of a state like $\alpha|b\rangle + \beta|c\rangle$ should indeed collapse it into either $|b\rangle$ or $|c\rangle$ with a probability respectively equal to $|\alpha|^2$ or $|\beta|^2$. This introduces, in any realist ontology, both indeterminism (God play with dice !) and non-locality (there are spooky action at a distance !). It also makes the theory incomplete insofar as the cut between what is and what is not described by the linear equation is not explained (when it is defined[2]). And the collapse, as I said above does not eliminate the parallel situations, it just confines them in a, sometimes accepted as such, unintelligible realm.

I also mention some physicists who believes there is neither collapse nor contagion. Some like Omnès, even accept the completeness of the description by the Schrödinger equation. Omnès is honest though, and ask us to abandon at this point the cartesian program for believing that [31]. Other, like Bohm or de Broglie, takes Schrödinger description as incomplete and shows that it can be completed by a non constructive (in the mathematical logician sense) and unknowable (in the physicist sense) set of initial particles positions from which a potential, acting non locally through the universal non collapsing wave described in the positions basis, will, in a determinist way, guide the

[2]This provides motivation for attributing the collapse to consciousness [42, 21]. The present approach can be seen as a no-collapse reconstruction of that idea from a classical theory of consciousness. This is going more toward a consciousness explanation of the quantum, than a quantum explanation of consciousness.

particles in a unique branch [10, 3]. Bohm's theory miss both the conceptual Occam Razor—the theory is heavy and difficult—and the ontological Occam Razor, because the potential need the entire wave function, i.e essentially the same many universes as Everett's one. Worst : we will see that from the computationalist point of view, Bohm's theory is unable to explain why we should expect staying ourself in that singularized branch. With computationalism, the very decoherence theory is relatively applicable in those empty branches, and this entails these "empty" particle free branches are noneless full of people like you and me, with similar observations on their worlds. If you propose to a bohmian quantum scientist to observe, relatively to the base $\{|1\rangle, |2\rangle\}$, a state like

$$\frac{1}{\sqrt{2}}|1\rangle + \frac{1}{\sqrt{2}}|2\rangle,$$

then, because he is bohmian, he will aknowledge the ontological existence of the resulting wave $\frac{1}{\sqrt{2}}|1\rangle|B_1\rangle + \frac{1}{\sqrt{2}}|2\rangle|B_2\rangle$, where B_i represents the quantum state of the bohmian scientist seeing i. If he was planning to measure some position, for exemple the position of his own body, the bohmian aknowledges that his *formal* doppelgänger will find definite results, like him, and explain them coherently with the decoherence theory (being, like him, a modern bohmian). In both Everett's and Bohm's view it can be predicted that the two bohmians will pretend to be observing particles and will assert that the other is a phantom, a bodyless zombie, someone without any *real* (with particles) body nor any consciousness[3]. But none can shown any evidence that he belongs to the branch whith the particles because the positions of these particles are necessarily *hidden*.

Bohm's theory is incompatible with computationalism because Schrödinger equation predicts that both B_1 and B_2 makes all the relevant computations for having a mind, with our without particles. But there is a lesson here: this illustrates we don't really need particles.

Let us show now that computationalism *per se* generalizes or even radicalizes Everett's monistic embedding of the subject into the object.

2 Arithmetical Realism

Let me put it in this way: *all* sufficiently realist interpretations of arithmetic accept the existence of parallel *situations*. This follows from the computationalist hypothesis asserting the existence of a level where we are Turing

[3]But still existing in some physical way through the empty wave which is still able in principle to influence the particle of his *unique* universe (there is no collapse).

Emulable[4]. This conclusion is not obvious nor is the notion of parallel situations clear in this setting. Actually such proof and clarification is one of the main result in my PhD thesis[5] [27]. The basic idea are simple, though, and, without going into much details I will try to convey briefly the main line of the argument.

The basic trick is the following : take a deep breath and introspect yourself, climbing all the way through your ancestors until you reach our common ancestor the amoeba [25, 26, 28]. I am used to propose some slower way and to define "being a mechanist" by "being someone accepting an artificial digital brain (in case of fatal brain disease for example)". It is easy to understand that in that case we are, like the amoeba, duplicable.

To make clear what is going on, it is necessary to distinguish two types of discourse: the first person and the third person discourses. The first person discourse is given by the result of experience/experiment which are written in a diary which belongs to the experimenter. It is important that he keeps his diary with him during the self-duplication experiment so that the diary is duplicated too in the experiment. The third person discourse is the discourse made by an external observer. Suppose a candidate goes through a self-duplication experiment, and that he believes in computationalism. He is scanned and annihilated at Brussels and reconstituted at both Washington and Moscow. Let us ask him the following question: where will you be after the experiment is done. He can answer: I will be in Washington *and* Moscow. Right, with *that* question, he can indeed use a third person discourse about himself. Let us ask him more cautiously the following question: where will you *feel yourself*, i.e. from your first person subjective point of view, after the experiment? More precisely: what will you note in your diary after the experiment is completed? In that case, if we assume both the computationalist hypothesis and if we assume that the experimenter has

[4]Precisely the computationalist hypothesis I use is the conjonction of 1) Arithmetical Realism (arithmetical truth are objective, independent of myself) ; 2) Church Thesis, (or better, the law of Post. See the formidable anticipation by Post 1922 [35, 9]). An anachronological modern version of Church's thesis is that all universal computer are equivalent in the sense that they defined the same set of computable functions on finite entities (the so-called partial recursive functions); 3) I am Turing emulable, in the sense that I can survive with an artificial digital "generalized" brain. A generalized brain is defined to be anything which is needed to emulate for my consciousness to proceed on. It could be the entire cosmos (in this case computationalism need the cosmos to be finitely describable). This made the reasoning very general. For making the reasoning easy I will suppose that the generalized brain is hot and circumscribed into our skull. The reader should convince himself that the reasoning still works with, for example, an unbounded quantum brain or any universal quantum turing machine [11].

[5]Loadable at http://iridia.ulb.ac.be/~marchal.

some minimal introspection abilities it is easy to understand he must answer 'I will feel myself in Washington *or* Moscow'. The fact is that he will not write in his diary something like 'I feel myself being both in Washington and Moscow'. In particular he will have a personal knowledge of Moscow (resp. Washington) and only an intellectual, 3-person knowledge, of the existence of his doppelgänger in Washington (resp. Moscow). And he can predict *that* very fact—that he will feel himself at one place—although he is unable to predict which one, and this shows that computationalism entails a strong first person indeterminacy, and this happens in the context of a strong third person determinism.

The next important point is the understanding that any quantification[6] made on that first person indeterminacy (1-indeterminacy for short) remains unchanged if we add delays of reconstitution. Such delays are not 1-observable. A sequence of simple, perhaps tedious, thought experiments leads toward a general invariance lemma:

> **Invariance Lemma:** The way you quantify 1-indeterminacy is independent of the (3-)time, the (3-)place and the (3-)real/virtual (or even purely arithmetical[7]) nature of the reconstitution[8].

The 1-experiencer, in self-multiplication experiments, cannot distinguish a real environment with a simulated one (at some level) nor can he observes any delays in the reconstitution, nor any space translation of these reconstitutions. To predict his own personal futur he must take into account all possible reconstitutions.

Now it is an easy, although quite astonishing, consequence of Church thesis that it exists a universal dovetailer program (UD), i.e. a program

[6]See the thesis by Nick Bostrom [6](also loadable at http://www.analytic.org/) for a general study of the self sampling assumption (SSA) principle. Such principle should make possible a reading of my approach in term of a universal-turing-machine-tropic principle, which generalizes the anthropic principle. Look at the archive of the Everything Mailing list for an entertaining discussion on similar matters: http://www.escribe.com/science/theory/.

[7]This last *arithmetical* point is not easy. It relies on a result showing the incompatibility between the physical supervenience and the computationalist hypothesis. That result has been obtained independently by myself and by T. Maudlin, [23, 29]. I don't really use that result in the present paper.

[8]The indeterminacy is pure 1-indeterminacy. Nevertheless, by duplicating entire population, the indeterminacy can be made third person 'verifiable' inside each multiplied population. This made the distinction between 3 and 1 point of view more relative than it could seem a priori. This is an important point to understand that our idealism is not a solipsism, but also to understand that some features of quantum mechanics, like phenomenological indeterminacy and non locality, can be seen as an a posteriori empirical confirmation of computationalism.

which is able to generate and simulate all other possible programs written in all possible programming languages, including all rational approximations of any universal unitary transformation, i.e. quantum programming languages.

Let us suppose that there is a concrete Universal Dovetailer UD running in our universe, and let us suppose, for the sake of the argument, that our universe is sufficiently robust for allowing the UD to never stop. Let us call UD* the infinite trace of the UD. It follows from the invariance lemma that the domain of 1-indeterminacy, for a candidate in computationnal state S, is the infinite set of all possible, relatively consistent, computational continuations of that state S appearing in UD*. In particular the UD will dovetail on any quantum algorithm, even on all solutions of all the Schrödinger equations associated with all Hamiltonians described in all programming languages[9]. But not only that: the UD generates all real numbers[10] as oracles, which a priori entails a high probability for appearances of typically highly non computational things, from white noise to hallucinatory experiences like the appearances of white rabbits[11]. Finding an (apparently) non computational object would not disprove computationalism, quite the contrary: there are apparently too much non computational objects.

Here we have "parallel situations" with a revenge : the Everett relative quantum states are, a priori, *particular* computational states. Would that last statement be proved (with respect to some measure), computationalism would be refuted (with respect to that measure).

With the computationalist hypothesis we have to define the indeterminacy on the set of all consistent and computationaly accessible extensions. So we must extract our continuous, computable and perhaps even linear anticipations from the quantification on that 1-indeterminacy. It looks like we have to find a measure on a set of relative computations to extract, hopefully

[9]See [24]. A similar idea appears under the form of *a Great Programmer* in[36] available as the quant-ph/9904050 at http://xxx.lanl.gov/. Despite key ideas Schmidhuber seems to ignore the distinction between first and third person point of view so that his way to eliminate the white rabbits seems to me not yet conclusive. Same remarks for Standish's approach [39] for the hunting of the white rabbit (loadable from http://parallel.hpc.unsw.edu.au/rks/pubs.html). Both Standish and Schmidhuber seems to believe we are inhabiting a well defined particular universe among all possible universes, but this has no meaning with the computational hypothesis as our work illustrates.

[10]Without being able to put them in a list. There is no contradiction with Cantor non enumerability theorem.

[11]Here we meet the inductive scepticism problem which appears with any kind of modal realism, see [19]. An entertaining discussion on white rabbits and wolf eating lamb in a similar context can been found in the archive of the discussion list at http://www.escribe.com/science/theory. For an advanced study on the original white rabbit look at [26, 8].

existing, lawful regularities.

Well that is the result: it is a problem, a challenge for those who takes mechanism seriously, by which I mean without abandoning the cartesian program. If there is a level such that we survive a functionnal substitution at that level, then we have to derive the laws of physics from the set of computations. We must explain why white rabbits or flying pigs are *so* rare.

Moreover, if quantum mechanics is the correct description of reality, we must derived it from computationalism. Phenomenological indeterminacy, under the form of many parallel situations, and non-locality are, as I have illustrate, not so difficult to derive. Interference, entanglement, and the classical part of quantum mechanics are much less obviously derivable. In the next section I illustrate a proposal inspired by physics then I come back to our 'pure' introspective computationalist road.

Of course I have used the hypothesis that there is a concrete, physical (whatever that means) implementation of a Universal Dovetailer. So perhaps I have only proved that mechanism entails our universe, perhaps unique, is too little or not sufficiently robust to process a significant part of the universal dovetailing. Actually that move is forbidden if we use the arithmetical part of the invariance lemma (see reference in footnote 5). We can also use some form of Occam razor instead, once we got evidence that a derivation of physics is plausible, which is what I illustrate in the two following sections.

3 The Hunting of the White Rabbit (I)

To illustrate the hunting of the white rabbit I propose here to search inspiration in physics. In the preceeding section I have showed that the foundation of physics and the foundation of psychology meet each other through mechanism, and it is not ridiculous to try to dig on both sides.

If there is an equation of the physical universe (like DeWitt Wheeler quantum relativistic equation) the UD will eventually generate it and will simulate its solutions. There is indeed evidence that our cosmos has a long computational history so perhaps such an equation *is* the solution. In our computationalist context this would only work if we are able to isolate that equation as the only one which in some sense supersedes, in term of relative measure of computational continuations, all the others. The invariance lemma indeed forces us to find a justification why our experience remains linked to that particular computation.

Let us ask a physicist where does the equations comes from. Let us ask him or her why force and acceleration are proportional (Newton's law). Why does 'F = ma' ?

A first typical physicist's answer is 'because it works'. Of course in our context such an answer does not work (we should have asked at once why does 'F = ma' works). A second, much more interesting physicist's answer is the following one: 'because nature tries to do the less'. Indeed it is possible to derive Newton's law from a minimisation of action principle. This is still a purely physicist answer as it is a reduction of a physical principle from another one. Although it provides us with deep information the fundamental questions remains. Why does nature need to act economically? Where does *action* comes from? Without forgetting still more fundamental questions like why are big phenomena reducible to a so little quantity of words (equations)? Like Wheeler[12] I doubt that such a question can be answered at all by an appeal to physical equations. The origin of empirical laws cannot be explained by an empirical law: that would entail an infinite regress. This does not mean that an idea inspired by physics cannot throw light on our hunting of white rabbits. We know that Newton's law can be reduced to an economy principle, so let us ask to a modern physicist where does such an economy principle comes from. Here is the most crazy, unexpected (but I guess correct) answer given by Feynman.

How does a particle find the shortest path from A to B? Well Feynman explains us that the particle just takes *all* the paths from A to B. But on each path from A to B a sine function is associated, and the probability that the particle goes from A to B is given by the square of the sum of all sine functions at B. On the many shorter paths the sinus function add constructively (because these paths have almost the same length) and on all extravaguant and lengthy paths the sinus add destructively because those path have arbitrary length so that the sinus phase are random, and this seems to explain, indeed, how nature manages to minimize action and why eventually F = ma. How clever! This is of course Feynman's path integrals formulation of quantum mechanics with his many parallel situations playing a so much decisive role.

The funny aspect of this explanation is that the extravagant path are eventually deleted not because these are rare but because these are numerous and extravagant (lengthy). Could it be that similarly the appearance of the white rabbit should be rare because, whatever its computational origin is, that origin will be superseded by its corresponding wave-like quantisation? Could it be that any classical deterministic explanation-program H is

[12]See Wheeler's 'Law without Law' [44].

superseded by a small variant (in number of bits) like e^{-iH}? Could it be that the Universal Dovetailer is supersed by the Quantum Universal Dovetailing he generates, and this through a link between relative complexity and phase randomization ?

4 The Hunting of the White Rabbit (II)

Let us come back to our pure introspective computationalist approach and let us try now to take seriously the needed distinction between third and first person description we have been obliged to introduce. To prevent our idealist approach from going toward solipsism, we need an 'objective theory of consciousness'. Clearly we need a serious axiomatic of consciousness and indexicals if we hope to extract the relevant Hilbertian measure on the set of computations. What follows is an attempt, with a modest result[13], in that direction. The basic idea here is to let a Universal Turing Machine to introspect itself and then just to interview it. As we will see we can do much more thanks to a formidable generalisation of Gödel's theorem due to Solovay [38]: indeed we will be able to interview not only the machine, but a sort of guardian angel associated with that machine, and which knows much more.

Like Lucas, or Penrose more recently, I agree one can make sense of the idea that Gödel's incompleteness theorems does apply to machines [22, 32]. In any case this becomes obvious if we restrict ourselves to the class of self-referentially sound Universal Turing Machines having enough arithmetical provability power to generate elementary arithmetical truth.

Unlike Lucas and Penrose, and just because I *postulate* mechanism at the start, I accept that, as far as we are sound machines, the incompleteness phenomena applies to us[14]. Like Webb I take Gödel's results as a protection of Church's thesis against "abusive" Penrose-like diagonalisation [43]. Church thesis itself makes the computationalist assumption at least well defined and consistent. In the spirit of Myhill I take the incompleteness theorems as the

[13]It is the second main result of my thesis cited above.

[14]The traditional use of Gödel's second incompleteness theorem for refuting mechanism is not logically valid. Nevertheless an important part of that proof can be reconstructed and it is possible to show that it only proves that *if* I am a machine then I cannot know which machine I am. Note that this is well illustrated with the self-duplication experiment where you cannot recognized yourself in the doppelgänger. A first, still invalid, reconstruction of Lucas' argument has been given by Benacerraf [1]. It has been corrected independently by a lot of people, notably Chihara, Reinhardt, Shapiro, and myself. See my thesis for details and references. Note that Penrose, in some subsequent books has corrected his Gödelian argument, but it seems that he doesn't take his correction into account in his philosophical argument.

first exact theorems in abstract or theoretical psychology [30]. Put in another way I just modelize third person communication by formal provability. And, inspired by Helmholtz's work on perception, I modelize consciousness by automatic, or instinctive, inductive inference of self-consistency[15].

Since Gödel 1931 there has been tremendous progress in the Gödelian study of provability. In an important 1955 paper Löb obtains a nice and non trivial generalisation of Gödel's second incompleteness theorem [20]. Let us write $\Box p$ for $provable(\ulcorner p \urcorner)$ where *provable* denotes Gödel's arithmetical provability predicate, and $\ulcorner p \urcorner$ denotes the Gödel number of a formula p [15]. We use $\Diamond p$ to abbreviate $\neg\Box\neg p$, where $\neg$ is the negation operator. $\Diamond p$ means that $\neg p$ is not provable (in the theory), which means that p is consistent (in the theory). The symbol $\top$ and $\bot$ denotes the propositional constant TRUE and FALSE. Gödel's second incompleteness theorem can then be written

$$\Diamond\top \rightarrow \neg\Box\Diamond\top$$

If the theory is consistent it cannot prove its consistency. This is obviously equivalent to $\Box(\Box\bot \rightarrow \bot) \rightarrow \Box\bot$. The formalisation of Löb's theorem is the more general formula L: $\Box(\Box A \rightarrow A) \rightarrow \Box A$. It is a theorem in Peano Arithmetic or in Zermelo Fraenkel set theory or in any sufficiently rich theory. It concerns our universal turing machine which has a sufficiently rich provability power.

Solovay has been able to show that the modal system G, having Löb's formula as the main axiom, (i.e. extending the basic Kripke modal system) formalizes soundly and completely formal provability[16]:

AXIOMS :	$\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$	K
	$\Box(\Box A \rightarrow A) \rightarrow \Box A)$	L
RULES :	$\frac{A\ ,\ A\rightarrow B}{B}$	MP
	$\frac{A}{\Box A}$	NEC

K is Kripke axiom. MP is the modus ponens rule and NEC is the necessitation rule. And then Solovay gives us a formidable gift. He shows that the

[15]Self-consciousness can then be modelized by phenomenologically less instinctive inference of self-consistency. This distinction is useful for keeping animal's consciousness meaningful. Theoretical computer science can then give a role to the inference of self-consistency, namely it gives self-speeding-up abilities relatively to the neighborhood. We can then speculate consciousness appears with the self-moving abilities of living being for self-moving needs the ability of anticipating neighborhoods.

[16]I suppose there is no variable in the scope of the provability predicate. In this case it has been shown that there is no complete formalisation. See the nice book by George Boolos for more information [5]. A gentle and recreative introduction to the modal provability logics is given by Raymond Smullyan [37].

following system, known as G*, having as axioms all the theorems of G, plus the reflexion formula $\Box A \rightarrow A$, but *without* the necessitation rule, formalize soundly and completely the set of all true provability (and consistency) statements, including those which are true and unprovable by the machine. So G* \ G captures all those true but unprovable statements. For example, the consistency sentence $\Diamond\top$ belongs to G* \ G. The guardian angel (G*) can communicate us that the machine is consistent, but the "machine itself" (G) cannot.

Unfortunately, although G and G* can aptly be considered as the logics of self-reference, the Gödelian formal provability concerns only third person discourses. Although through the Gödel's numbering technic the machine communicates genuily about itself, the machine does so in an impersonal way: it is really third person self-reference.

How to get the first person? That is an hard matter, so let us climb on the shoulders of giants.

Plato, in the mouth of Thaetetus[33], proposes us to define *knowing* p by *justifying* p which I take as *provable* ($\ulcorner p \urcorner$). Socrate argues that some justification can be wrong. So Thaetetus proposes to define *knowing* p by *justifying* p with p true, *by definition*. This is a nice idea. Unfortunately we cannot translate it in the language of our *Löbian* machine. By Tarski theorem, we cannot define a truth predicate $true(\ulcorner p \urcorner)$, so we cannot express in the language of the machine "$provable(\ulcorner p \urcorner)$ & $true(\ulcorner p \urcorner)$".

But here is a trick, we can represent *knowing* p by *provable* ($\ulcorner p \urcorner$) & p. And when you do that, like Boolos and Goldblatt in an independent way [4, 17], you get two new modal systems S4Grz and S4Grz*, which happen to be equal: S4Grz = S4Grz*, i.e.the guardian angel doesn't add anything, i.e. from the point of view of the subject, truth is provability, provability is truth. This gives us a kind of very solipsistic first person theory akin to Brouwer's consciousness theory [7]. In particular Goldblatt has derived from S4Grz an arithmetical interpretation of intuitionistic logic [17, 18].

Unfortunately this *subject* is a little too much solipsistic, and besides, we are not searching a knowing subject, but a sensible observer. We remember that our thought experience have shown that, through computationalism, the observer's experiences at some stage are defined by all their consistent and accessible (by the universal dovetailer) extensions or continuations, relatively to that stage. So let us weaken truth by possibility or consistency. This is a natural move in any indexical approach of actuality, like Everett's one. It also fit nicely with our (re)definition of physics as an uncertainty-quantification theory where the 'probabilities' are defined on the domain of the relative consistent computational extensions. So let us defined our first person observation, not as $\Box p$ & p, but as $\Box p$ & $\Diamond p$. We get a couple of

logic[17], which I call Z and Z*, proving the formula $\Box p \rightarrow \Diamond p$, and this is nice because it is (obviously) a welcomed formula for the modelisation of certainty in the modal approach of probability (see also [14]).

We must still ask the universal machine to take into account *computationalism*, I mean we must not forget that the extensions must not only be consistent, but must also be accessible by the universal dovetailer. In logic an arithmetical formula is said to be Σ_1 if it can be put in the form $\exists x P(x)$ with P (algorithmically) decidable. If such formula are true then they are provable, they are verifiable and they correspond to the leaves of the universal dovetailer[18]. So our phenomenological physics is given by the weakening of Thaetetus' idea $\Box p \ \& \ \Diamond p$, with p verifiable, UD-accessible, or simply Σ_1.

Löbian machines can prove their own Σ_1-completeness in the sense that they can proof $p \rightarrow \Box p$ for all Σ_1 sentence p, and A. Visser has proved that $G + p \rightarrow \Box p$, gives a complete and sound axiom system for the provability when the arithmetical interpretation of atomic formula is restricted on the Σ_1 formula [41].

Visser result gives us a theorem prover for such logics and this gives us two new logics Z_1 and Z_1*. Our result is that Z_1* proves

$$p \rightarrow \Box \Diamond p,$$

and this is *very* nice, because, thanks to an important work by Goldblatt again [16], this gives us a way toward an arithmetical interpretation of quantum logics and quantum probabilities. It is the gap between Z_1 and Z_1* which makes possible to differentiate the notions of communicable observable, which I conjecture correspond to physical measurement, and uncommunicable observable which could be use for a theory of *qualia*.

Before we open the champagne bottle, I must mention we loose, with the Z logics, the necessitation rule. Because of that we loose the facilities of Kripke world-semantics, and, in fact, at this stage it can be argued that we loose *all* the universes. The physical predicates become psychological. We are really going toward a many-minds/no worlds interpretation of quantum mechanics and arithmetics. Technically it means also we don't really get the Birkhoff-von Neumann quantum logic [2], but a variant of it. Let us just hope

[17]From now on, a logic is considered as a set of formula (the theorems) and *not* as an axiomatised presentation. Actually none of the Z(*) logics which follow have been axiomatized, nor even proved to be axiomatizable. These are open problems. Nevertheless, thanks to Solovay theorems, these logics are decidable, and it is easy to write demonstrator algorithm for each one.

[18]In a nice introduction to model theory, Bruno Poizat gives the following form to Church's thesis: a function is computable if it is Σ_1. He means iff its graph, i.e. the set of couples $(x, f(x))$ can be defined by a Σ_1 formula [34].

such logics will help us to find our way in the labyrinth of quantum logics [40]. Nevertheless our logics are very constrained, by the Gödelian nuances, so that there is room for proving the unicity of measure on the set of relative computational continuations, and for extracting the quantum formalism and its classical limits.

References

[1] P. Benacerraf. God, the Devil, and Gödel. *The monist*, 51:9–32, 1967.

[2] G. Birkhoff and J. von Neumann. The Logic of Quantum Mechanics. *Annals of Mathematics*, 37(4):823–843, 1936.

[3] D. Bohm and B. J. Hiley. *The undivided universe*. Routledge, London and New York, 1993.

[4] G. Boolos. On Systems of Modal Logic with Provability Interpretations. *Theoria*, 46(1):7–18, 1980.

[5] G. Boolos. *The Logic of Provability*. Cambridge University Press, Cambridge, 1993.

[6] N. Bostrom. *Observational Selection Effects and Probability*. PhD thesis, London School of Economics, 2000.

[7] L. E. J. Brouwer. Consciousness, philosophy and mathematics. In P. Benacerraf and Putnam H., editors, *Philosophy of Mathematics*, pages 90–96. Cambridge University Press, Cambridge, second edition, 1983. première édition chez Prentice-Hall 1964.

[8] L. Carroll. *Alice in Wonderland*. MacMillan, London, 1865.

[9] M. Davis, editor. *The Undecidable*. Raven Press, Hewlett, New York, 1965.

[10] L. de Broglie. *La théorie de la mesure en mécanique ondulatoire*. Gauthier-Villar, Paris, 1957.

[11] D. Deutsch. Quantum theory, the Church-Turing principle and the universal quantum computer. *Proc. R. Soc. Ac.*, 400:97–117, 1985.

[12] H. Everett III. "Relative state" formulation of quantum mechanics. *Review of Modern Physics*, 9(3):454–462, 1957. Aussi dans DeWitt et Graham 1973.

[13] H. Everett III. The theory of the universal wave functions. In B. DeWitt and N. Graham, editors, *The Many-Worlds Interpretation of Quantum Mechanics*, pages 3–140. Princeton University Press, Princeton, New Jersey, 1973.

[14] M. Fattorosi-Barnaba and G. Amati. Modal Operators with Probabilistic Interpretations i. *Studia Logica*, XLVI(4):383–393, 1987.

[15] K. Gödel. Über formal unentscheidbare sätze der principia mathematica und verwandter systeme i. *Monatsh., Math. Phys.*, 38:173–198, 1931. Traduction américaine dans Davis 1965, page 5+.

[16] R. I. Goldblatt. Semantic Analysis of Orthologic. *Journal of Philosophical Logic*, 3:19–35, 1974. Also in Goldblatt 1993, page 81–97.

[17] R. I. Goldblatt. Arithmetical Necessity, Provability and Intuitionistic Logic. *Theoria*, 44:38–46, 1978. Aussi dans Goldblatt 1993, page 105–112.

[18] R. I. Goldblatt. *Mathematics of Modality*. CSLI Lectures Notes, Stanford California, 1993.

[19] David Lewis. *On the Plurality of Worlds*. Basil Blackwell, Oxford, 1986.

[20] M. H. Löb. Solution of a problem of Leon Henkin. *Journal of Symbolic Logic*, 20:115–118, 1955.

[21] F. London and E. Bauer. *La théorie de l'observation en mécanique quantique*. Hermann et Cie, Paris, 1939.

[22] J. R. Lucas. Minds, Machines and Gödel. *Philosophy*, 36:112–127, 1961.

[23] B. Marchal. Informatique théorique et philosophie de l'esprit. In *Actes du 3ème colloque international de l'ARC*, pages 193–227, Toulouse, 1988.

[24] B. Marchal. Mechanism and personal identity. In M. De Glas and D. Gabbay, editors, *Proceedings of WOCFAI 91*, pages 335–345, Paris, 1991. Angkor.

[25] B. Marchal. Amoeba, planaria, and dreaming machines. In P. Bourgine and F. J. Varela, editors, *Artificial Life, towards a practice of autonomous systems, ECAL 91*, pages 429–440. MIT Press, 1992.

[26] B. Marchal. Conscience et Mécanisme. Technical Report TR/IRIDIA/95, Brussels University, 1995.

[27] B. Marchal. *Calculabilité, Physique et Cognition*. PhD thesis, Université de Lille, Département d'informatique, Lille, France, 1998.

[28] B. Marchal. *Le secret de l'amibe*. Partage du Savoir. Grasset, Paris, 2001. Forthcoming.

[29] T. Maudlin. Computation and Consciousness. *The Journal of Philosophy*, pages 407–432, 1989.

[30] J. Myhill. Some philosophical implications of mathematical logic. *The review of Metaphysics*, VI(2), 1952.

[31] R. Omnès. *Quantum Philosophy*. Princeton University Press, 1999.

[32] R. Penrose. *The Emperor's New Mind*. Oxford University Press, Oxford, 1989.

[33] Platon. *Théétète ou de la science*, pages 83–192. Oeuvre de la pléiade. Editions Gallimard, Paris, 1950.

[34] B. Poizat. *A Course in Model Theory*. Springer, New-york, 2000.

[35] E. Post. Absolutely unsolvable problems and relatively undecidable propositions : Account of an anticipation. In Davis [9], pages 338–433. 1922.

[36] J. Schmidhuber. A Computer Scientist's View of Life, the Universe, and Everything. In C. Freksa, editor, *Foundations of Computer Science: Potential-Theory-Cognition*, pages 201–208. Lectures Notes in Computer Science, Springer, 1997.

[37] R. Smullyan. *Forever Undecided*. Knopf, New York, 1987.

[38] R. M. Solovay. Provability Interpretation of Modal Logic. *Journal of Mathematics*, 25:287–304, 1976.

[39] R. K. Standish. Why Occam's Razor. Submitted to Annals of Physics.

[40] B. C. van Fraassen. The labyrinth of quantum logic. In R. Cohen and M. Wartosky, editors, *Boston Studies of Philosophy of Sciences*, volume 13, pages 224–254. Reidel, Dordrecht, 1974.

[41] A. Visser. *Aspects of Diagonalization and Provability*. PhD thesis, University of Utrecht, Department of Philosophy, The Nederland, 1985.

[42] J. von Neumann. *Mathematical Foundations of Quantum Mechanics*. Princeton University Press, NJ, 1955. Edition originale allemande: 1932.

[43] J. C. Webb. *Mechanism, Mentalism and Metamathematics: An essay on Finitism*. D. Reidel Publishing Company, Dordrecht, Holland, 1980.

[44] J. A. Wheeler. Law without Law. In P. Medawar and Shelley J., editors, *Structure in Science and Art*, pages 132–154. Elsevier North-Holland, Amsterdam, 1980.