CONVERGENCE AND FREE-WILL

CRISTI STOICA

Abstract. If our mind is just an algorithm running on a flesh hardware, then it seems that there is no place for the free-will. An algorithm decides everything based on deterministic computations, or on random inputs, but neither inevitability nor pure hazard is free choice. Hopefully, some day, Science will be able to understand, monitor and simulate all the mind processes. Even then, it will still be a possibility for the free-will to exist, based on the convergence of the initial data. I propose a crucial experiment to test this hypothesis.

Contents

1. Is there a ghost in the machine?

The mind-body problem is of a fundamental importance for the human beings. We all want, or need, to know more about ourselves. The mind and matter are very real things, and cannot be dismissed that easily. It is no wonder that, at the beginning, many philosophers considered that they both exist, and cannot be reduced one to the other. In the same time, some thinkers, mostly in orient, proposed that there is only the mind, and the material world is an illusion. Yet, Science evolved in such a manner that provided a physical basis for more and more phenomena, which initially were though to be irreducible to matter. This progress led today to such a wide mapping of the phenomena related to mind, consciousness, thinking, feelings, that it seems that there is only one percent remained unexplained, and it is only a matter of time that we will soon

Email: holotronix@gmail.com.

provide an explanation of all the mind phenomena. Thus, everything will be reduced to materialist explanations. But, it would not be the first time scientists have the illusion that little remained to be explained. For example, at the end of the XIX-th century, it was believed that Physics is almost closed, then, in only several years, the Theory of Relativity and Quantum Mechanics opened new realms, much wider than the classical Physics.

Well, for some scientists may be clear that the missing "percent" in our explanations about mind will be explained soon, as usual, by reducing it to physical phenomena. Others may think that what remained unexplained is really important and irreducible. For example, it is difficult to conceive how it will be provided an explanation of the acute feeling we all have, that we really experience our lives, and we are not just a software running on an organic hardware. And it is again difficult to admit that our bodies just follow the laws of Physics, when we are so aware that in fact we are driving them, obviously with a purpose.

This is why I think that we cannot draw yet a definitive conclusion in the mind-body problem. I will not add many arguments for one viewpoint or another. Instead, I will propose a localization of the problem, such that we can formulate it in a way that emphasizes better a key aspect. Then, starting from this key problem, I will propose a crucial experiment that may provide an objective resolution of the mind-body problem.

2. **Minds in a physical world**

The body related part of the problem involves inevitably the laws of Physics. We don't know yet everything about the physical universe in which we live. Our best theories, General Relativity and The Standard Model of Particle Physics, can explain a large part of the physical world, but still remains something unexplained. For example, we just don't know yet how to combine these two theories, despites of the progresses made in String Theory and Quantum Gravity (which are very different than the theories they try to unify). Even if we believe that only 1% remains unexplained, we can at most hope that it will be dissolved in what we know so far. In fact, there are unexplained things in Quantum Mechanics and Relativity too. Even when we will have a Theory of Everything in Physics, corroborating what we know so far, there will always exist the possibility to be contradicted by future experiments.

Although we don't know which theory will win, and how our final explanation will look like, we can make some general assumptions, which are respected by probably all the theories proposed so far in Physics. As I described in [1], in most our theories and attempts, the spacetime can be represented mathematically as a topological space (including the discrete cases). The laws of Physics can be interpreted as imposing constrains on the fields allowed on the spacetime. We can therefore identify the laws with sections of a topological bundle over the spacetime. There are, of course, more than one field, but can we always combine them with ease in a single field, by operations with bundles. In order to distinguish the local laws of Physics from the global laws, we can consider instead of the global sections of the bundle, a sheaf over the spacetime, with the requirement that the solutions will always be global sections of that sheaf. I will not

repeat here the details, they can be found in the cited article. I will only mention that, even if we want to model theories that cannot, in principle, be represented as sheaves over topological spaces, we can consider the spacetime to be a category, and the law sheaf a contravariant functor on it. This is based on the fact that any topological space can be viewed as a category.

The mathematical structure constructed in this way is named World. The analysis presented there can make us an idea of what we can expect from our theories, independent of which of them will be accepted as the description of our world. Even more, we can see that the World Theory allows an inclusion of the emergent phenomena, as extra conditions that simply refine the law sheaf. Metaphorically, some emergent phenomena can be thought as software, running on the hardware formed by the matter fields.

But can we identify the mental processes with a software, implementable in terms of laws of Physics? I don't know. What I can say is that a lot of progresses are made, in understanding the mind in terms of neurons and brain. Will, in time, the (more or less) tiny part remained unexplained, be explained in terms of matter? I showed in [2] that the Turing test [3] can easily be passed by an algorithm, but this may not be that significant. The main challenge is to understand how is made that software/hardware combination.

And now we come to the main problem. Let's be optimistic and suppose that, some day in the future, we will be able to monitor the brain activity, and to point precisely how is constructed the mind, and how it works, as a software. The question many of us have, is this: *will that software really be a mind?*. Our instincts may say that there is no way that we are just such a software. Our experience, which accustomed us with software able to do unimaginable task, saids that it may be. But is it? Our smart, or genius, cell phone that we will use at that time, will cry to charge its batteries, but will this mean that it really feels hunger?

3. The problem of free-will

3.1. What do we think we have, and a machine cannot have?

Although it is very possible to simulate a human mind in such a way that we can't distinguish the simulation from an original mind, will this mean that really there is no difference?

An operationalist view will say that, because we can't find any objective difference, there is no one.

Some differences are suspected to exist. I believe that the more important that were raised are the following ones:

Problem 1. Real human minds can have a strong feeling of free will.

Problem 2. I, as a human, have a very clear feeling that I exist, that I am a subject experiencing the processes of my mind.

Somebody may object that I let aside the important issue of feelings, love, generosity etc. This may be a point, but I will not discuss them here. I only want to say that a

large range of feelings has been showed to be accompanied by brain processes that can be monitored, and we can use these processes to make predictions about the feelings. It is easy to believe that this part of the mind activity will be describable by a mind software. What is more subtle is the feeling that we really are something having those feelings, hence the problem 2. If we monitor a brain in all details, and see the processes corresponding to a given feeling, is this a guarantee that there is somebody "in there" really having the feelings?

I believe that the two problems are strongly related, but I will focus in this article on the free-will problem.

3.2. **The problem of free-will**

The problem of free will (Problem 1) has been related to the problem of determinism. Before the advent of Quantum Mechanics, the laws of Physics were deterministic. This feature resides in their expression as partial differential equation (PDE), subject to a set of initial conditions. Knowing the initial conditions, the solution is uniquely determined. Even areas of Physics which initially were considered to be subject of randomness, were explained by statistics of systems made of large numbers of particles or molecules.

The determinism seemed for many to be a menace at the address of the free-will, because, if everything is determined by some initial conditions in a very distant past, it is too late for us to make any decision which is not predetermined.

Quantum Mechanics introduced a probability, which seemed irreducible to deterministic laws. For physicists like Einstein[4], de Broglie, Bohm[5, 6, 7], Vigier, Schrödinger, to accept this probabilistic character was an abdication from a fundamental principle of science, the search for a more complete explanation. Bohr, Heisenberg, Born, on the other hand, considered that these probabilities are irreducible. For many this was very attractive feature, since it is commonly believed that it allows, if not guarantees, the free-will. There is a common belief that the determinism is incompatible with free-will[8, 9].

3.3. **Is the free-will allowed by Quantum Mechanics?**

J.S. Bell showed in [10] that the idea of pre-programming some possible hidden variables to simulate the randomness of Quantum Mechanics can be tested experimentally. Such experiments were carried out since then [11, 12, 13], confirming that there is no room for pre-programming. This seemed to rule out any possibility for the deterministic view.

On the other hand, the idea of wave function collapse, which is the mechanism that introduces the probabilities, has some ugly properties, suggesting that it would be desirable to find a better explanation. Such explanations were provided during the time, but the main problems remained. Even the Many Worlds interpretation [14, 15, 16, 17, 18, 19], although claiming to solve this problem, it lets untouched the problems for the actual beings living in the actual worlds supposed to split [20, 21]. The decoherence approach [22, 23, 24, 25, 26, 27, 28, 29, 30] seems to be the solution for a part of the quantum measurement problem and the emergence of classical from quantum, but there is an important part still escaping.

In a paper named "Smooth Quantum Mechanics " [21], I showed a possible solution to the apparent discontinuity occurring during the wave function collapse. Instead of accepting that the Schrödinger equation is violated from time to time, by a discontinuous jump, I consider another explanation. There is a feature which is present even in standard Quantum Mechanics, namely, what we choose to observe, affects in a strange way the past![31, 32, 33, 34, 35, 20]. The delayed choice made by the observer partially determines the solution of the Schrödinger equation, in an irreducible manner. I use this feature, combined with the entanglement between the observed system and the measurement device used for its preparation, to explain how is it possible that any observation finds the system in an eigenstate, without the need of discontinuities in the evolution. This can be done by allowing the initial conditions of the observed quantum system plus the preparation device to be delayed. The *delayed initial conditions* mechanism shows that it is never too late to have a decision concerning the initial conditions of the Universe (as long as they are not established yet). This mechanism reintroduces the determinism in Quantum Mechanics, in a way that is simpler than in the standard approach, avoiding the problems of the wave function collapse. This kind of determinism is compatible with the free-will at the same extent as the indeterministic interpretation.

Whether or not this mechanism will be confirmed, Quantum Mechanics indeed can allow the observer to make a choice. Either if this is done by discontinuities in the evolution (which require a restatement of the initial conditions), or by the Smooth Quantum Mechanics, the observer (that's us) has a word to say about the evolution of the quantum system. I will call this feature, in the deterministic or in the indeterministic form, the *quantum freedom*. These two possibilities seem also to be the only ones allowed by the World Theory. But, allowing the free-will, is not to say that it is guaranteed.

3.4. **Is there enough room for the free-will in the brain processes?**

3.4.1. **Quantum mind**

Various ways to use the quantum freedom in a quantum mind model were proposed by various scientists, for example Henry Stapp [36, 37], Penrose and Hameroff [38, 39, 40, 41]. Tegmark [42, 43] showed by calculations that the decoherence happens too fast, and the brain cannot make use of the quantum behavior, at least in Penrose's theory.

The progresses in explaining how the brain works are mostly based on classical physics; we don't have any clue of quantum phenomena relevant to the brain functioning. This seems to imply that there is no room for making use of the quantum freedom, or at least the quantum phenomena will be insignificant for the mind processes. But is the quantum mind the only way?

3.4.2. **Can a computer algorithm have free-will?**

Let's first explore the following experiment. Suppose we create a software simulation of the mind, and then we copy that software, and run it on two identical computers. Then, suppose we input identical data, in order to test the reaction of that software. It's state will change, as a result of the interaction, perhaps it will also provide some output.

Will the two copies evolve identically, under the same input data? If that software implements a deterministic algorithm, it will evolve deterministic, hence identical on the two computers. If the algorithm is indeterministic, there will be some randomness present. That randomness is provided by the computer's random generator, and it depends on it. This is not a genuine free-will, because it is subject to the randomness of the random generator of the computer, and not of a free choice. On the other hand, what is a free choice? When an algorithm chooses something, it can do this either by a computation, or by using the randomness, or a combination. But none of these is really a free choice. Neither inevitability, nor pure hazard, is free choice.

If human mind is an algorithm, then the computer is the brain, or more generally, a part of the physical world. If the physical laws allow random input, as it seems to result from the quantum freedom, in both its versions, then this is like the random generator of the computer. Would this mean that there is no possibility for the free-will? Is our feeling that we can make a choice, nothing but an illusion of an algorithm, which in fact is tied to deterministic computation combined with random input?

It seems that the only possibilities are deterministic computation and random choice, and none of them can be identified with the free-will.

3.4.3. **Can we draw a conclusion?**

These two arguments suggest that there is no room for the free-will. But the mind-body problem is much closed to us, as human beings, and it deserves more attention; we cannot say too easy that we understand it well enough, and draw a quick conclusion.

Whether or not there is free-will, it is not yet clear. We don't know yet everything about quantum, or about mind. In the same way the 1% which it was believed to remain uncovered by the science at the end of the XIX-th century exploded in few years into the incredible worlds of quanta and relativity, we may expect big surprises.

As we shall see, it is too early to conclude.

4. The free-will test

4.1. Preconditions for the test

For the reasons exposed so far, I propose the following solution to the problem. First, we have to be a little more patient, until we understand better the physical level, and the way our minds work. I don't mean to say that we should stop trying to find a solution, by contrary. I recommend to work hardly in understanding better the matter and the mind, but to refrain to draw any definitive conclusion in the free-will problem, until we are able to make the following experiment, which can be crucial, and which depends on our progresses in the areas mentioned above.

The state of art in Science, when we will be able to perform the experiment, will have to fulfill the following conditions:

(1) We understand how the human brain works, in every detail.
(2) We know precisely how Physics explains all the details in the brain's functioning.

(3) We know how to relate each process taking place in the mind, with its correspondent process in the brain.
(4) We are able to simulate by a software the human mind, not only to be able to pass the Turing test, but also to contain simulations for each of the mind processes.

Are these conditions too strong to be ever fulfilled? Assuming that they will be sometime in the future realized, wouldn't this mean that we know everything about the human mind? Wouldn't this mean that the mind is just a software running on our brains, just a foam on the ocean of matter?

In understanding our minds, we will have to go on the scientific route. Currently, our science searches for materialistic explanations of everything. Perhaps some day we will construct a materialistic explanation for all phenomena, or, by contrary, we will need to expand the Science, such that it contains other than the usual matter phenomena. We don't know yet. But, since currently a lot of scientists already try to realize the conditions above, we have to advance together in this direction.

Assuming that the conditions above are fulfilled, many scientists will be satisfied and consider that we know everything about the mind, which is apparently only a manifestation of matter. Even if a human has the feeling that experiences something more, we will say that this feeling is also contained in the explanation we have about the brain's functionality.

I will show that, when these conditions are realized, something more needs to be done. There is still a possibility for our minds to be something more than just interacting matter, and this possibility needs to be confirmed, or infirmed, by an experiment.

4.2. **The "intention beyond matter" hypothesis**

Assuming the conditions from §4.1 are fulfilled, we can say that the human mind works like an algorithm. This algorithm can use some input, quantum for example, which allows it to be an indeterministic algorithm. We may believe that this is the end, and we are just algorithms.

What if the material mind is controlled from "outside", by an immaterial intention? For example, when a human plays a computer role-playing game, there is an algorithm representing the character in the game. The algorithm may use some random input, to act more unpredictable. But it is the human that also inputs data, and controls the character's actions. From the point of view of the algorithm, this looks like random input, but if the player is good, we can observe that the actions seem to have a definite intention. The inputs converge towards a purpose. Likewise, it is always possible that the material mind is just a character in a role-playing game, and the real "I" of that mind is something outside the material world. This immaterial "I" inputs, by using the quantum freedom, data into the material world, to control the decisions of the algorithm that we consider to be our minds.

This hypothesis has a big problem. It explains something from our world, by assuming something outside the world, which is less plausible, or at least less understood. Hence, we can consider that it does not help us too much.

This is indeed true. But, if we really want to make sure that there is nothing more about the mind, we will need to test this hypothesis. A skeptic will say that it is useless. I want to say that we can make a test to accept/reject the *intention beyond matter* hypothesis, and that it worth do it. It is important to do it, because it is an important piece to understand our minds. If it fails, then the skeptics can say that they have the proof. If, by any chance, succeeds, then there is no need to explain why it worth. In both cases, the way we understand our minds will be changed.

4.3. **The free-will test**

Imagine the role-playing game software described above. Assume that there is a part of that software, that have information about the behavior of a character of the game. Based on that information, this part may be able to tell whether the input controlling that character is purely random, or that it has a definite intention. What it need to do is to calculate the average points that the character can obtain to the game, if the input is random. Then, it calculates, based on the real score obtained by the character, how special must be the input. If the score is higher, then the input is directed towards a purpose.

When our knowledge will satisfy the conditions from §4.1, we will be able to tell whether the seemingly random inputs in our minds reveal an intention. These inputs are those provided by the quantum freedom (indeterministic or deterministic, depends upon the theory). The inputs being not determined by the physical fields in the world, we can say that they are from outside the world. They can be either random, or with an intention. We will have to calculate how many of the possible random inputs look like being with intention. We expect that this fraction is very tiny. Then, if by monitoring a human brain in action, we observe that its actions attain purposes that, by random inputs, are much less probable to be attained, we can conclude that it's activity is guided by an intention. And this intention is outside the physical world. If, on the other hand, the inputs look really random, then we have no reason to admit another intention, beyond the physical world.

The test consists in looking for improbable "coincidences" in the initial conditions of the world, that "conspire" into leading us to make choices that reveal intention. I name this type of phenomenon *convergence* of the initial conditions – it is like "God's dices" selecting the initial conditions are "loaded".

If the test reveals that the world is affected, via initial conditions, in such a way, we can conclude that we have a "component" which is irreducible to the physical laws, but which affects their initial data. This influence has an atemporal, or trans-temporal behavior, because it acts to initial conditions that are expected to be fixed long before our choice.

If the result of the free-will test proposed here will be positive, we can conclude that the answer to the problem 2 is that, indeed, beneath the brain processes that can be monitored by physical means, there is something more. If the result is negative, we can conclude that, if there is indeed a subject experiencing the phenomena, this subject must be totally determined by the physical laws. From scientific point of view, we will not be able to find differences between a human mind, and an artificial algorithm reproducing

exactly the processes and behavior of a human mind. It will remain only a matter of belief to consider that there exist an important difference, taking the form of a subject experiencing the mind processes, between the human mind and the artificial mind.

References

[1] Cristi Stoica. World theory. *PhilSci Archive*, 2008.

[2] Cristi Stoica. Turing test, easy to pass; human mind, hard to understand. *PhilSci Archive*, 2008.

[3] A.M. Turing. Computing machinery and intelligence. *Mind*, (59):433–460, 1950.

[4] A Einstein, B Podolsky, and N Rosen. Can quantum-mechanical description of physical reality be considered complete?, physical review 47:777–780, 1935.

[5] D Bohm. The paradox of einstein, rosen, and podolsky. *Quantum Th.*, pages 611–623, 1951.

[6] D Bohm. A suggested interpretation of quantum mechanics in terms of "hidden variables", i and ii. *Phys. Rev*, (85), 1952.

[7] D Bohm and B Hiley. The undivided universe: an ontological interpretation of quantum mechanics. routledge and kegan, 1993.

[8] John Conway and Simon Kochen. The free will theorem. *Found. Phys.*, (36):1441–1473, 2006.

[9] John Conway and Simon Kochen. The strong free will theorem.

[10] J S Bell. On the einstein-podolsky-rosen paradox. 1964.

[11] J F Clauser, M A Horne, A Shimony, and R A Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, (23), 1969.

[12] J F Clauser and A Shimony. Bell's theorem : Experimental tests and implications. In *Reports in the Progress of Physics*, 1978.

[13] A Aspect, J Dalibard, and G Roger. Experimental test of bell's inequalities using time-varying analyzers. *Physical Review Letters*, (49), 1982.

[14] Hugh Everett. "relative state" formulation of quantum mechanics. *Rev. Mod. Phys.*, 29(3):454–462, Jul 1957.

[15] H Everett. The theory of the universal wave function. In *in The Many-Worlds Hypothesis of Quantum Mechanics*, pages 3–137. Press, 1973.

[16] B S de Witt. The many-universes interpretation of quantum mechanics, 1971.

[17] B S de Witt and N Graham. The many-worlds interpretation of quantum mechanics, 1973.

[18] D Deutsch. *Int. J. Theor. Phys.*, (24):1–41, 1985.

[19] D Deutsch. Quantum theory of probability and decisions. *Proc. Roy. Soc. London*, (A455):3129, 1999.

[20] Cristi Stoica. A direct interpretation of quantum mechanics. *PhilSci Archive*, 2008.

[21] Cristi Stoica. Smooth quantum mechanics. *PhilSci Archive*, 2008.

[22] R Omnés. Logical reformulation of quantum mechanics. *Journal of Statistical Physics*, (53):893–932, 1988.

[23] R Omnés. Consistent interpretations of quantum mechanics. *Rev. Mod. Phys.*, (64):339–382, 1992.

[24] R Omnés. The interpretation of quantum mechanics, 1994.

[25] H. D. Zeh. What is achieved by decoherence? 1996.

[26] W. H. Zurek. Decoherence, einselection, and the existential interpretation (the rough guide). *Phil. Trans. Roy. Soc. Lond.*, A356:1793, 1998.

[27] W. H. Zurek. Decoherence, einselection, and the quantum origins of the classical. *Rev. Mod. Phys.*, 75:715, 2002.

[28] W. H. Zurek. Environment-assisted invariance, entanglement, and probabilities in quantum physics. *Phys. Rev. Lett.*, 90, 2003.

[29] W. H. Zurek. Quantum darwinism and envariance. 2003.

[30] W. H. Zurek. Probabilities from envariance. 2004.

[31] K F Weiszäcker. *Zeit. F. Phys.*, (70):114, 1931.

[32] Paul Arthur Schilpp Ed. *Albert Einstein : Philosopher-Scientist. The library of living philosophers, vol VII.* Cambridge University Press, 1949.
[33] J Wheeler. Include the observer in the wave function. In *Quantum Mechanics a Half Century Later*, pages 1–18, 1977.
[34] J A Wheeler. The 'past' and the 'delayed-choice' experiment. In *In A.R. Marlow, ed.,* Mathematical Foundations of Quantum Theory, page 30, 1978.
[35] J A Wheeler and W H Zurek. Quantum theory and measurement, 1983.
[36] H. P. Stapp. *Mind, matter, and quantum mechanics.* Berlin: Springer, 2003.
[37] H. P. Stapp. *Mindful universe: Quantum mechanics and the participating observer.* Berlin: Springer, 2007.
[38] Roger Penrose. *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics.* Oxford University Press, Inc., New York, NY, USA, 1989.
[39] Roger Penrose. *Shadows of the Mind : A Search for the Missing Science of Consciousness.* Oxford University Press, Inc., New York, NY, USA, 1994.
[40] Stuart Hameroff and Roger Penrose. Orchestrated reduction of quantum coherence in brain microtubules: a model for consciousness. *Math. Comput. Simul.*, 40(3-4):453–480, 1996.
[41] Roger Penrose. *The Large, the Small and the Human Mind.* Cambridge University Press, Inc., New York, NY, USA, 1997.
[42] Max Tegmark. The importance of quantum decoherence in brain processes. *Phys. Rev. E*, (61):4194–4206, 2000.
[43] Max Tegmark. Why the brain is probably not a quantum computer. *Information Sciences*, (128):4194–4206, 2000.