Wigner's weird friends

Pablo Garcia Tello, Leuven, Belgium, September 2009

Arguably the physicist Eugene Wigner can be considered as the one responsible for granting the status of *"respectful"* to the debate concerning the relationship between quantum theory and consciousness among the scientific community [1].

When the province of physical theory was extended to encompass microscopic phenomena, through the creation of quantum mechanics, the concept of consciousness came to the fore again: it was not possible to formulate the laws of quantum mechanics without reference to the consciousness.

The words above, taken from his 1961 paper "Remarks on the Mind-Body Problem", served him as an introduction for considering the fundamental role played by consciousness in quantum theory. His opinions were brilliantly exposed with the aid of a nowadays classical thought experiment involving a hypothetical friend to be known for ever as *Wigner's friend*. Since Wigner's landmark paper audacious individuals have entered this *quantum vs. consciousness* uncharted land trying to make sense of a problem that in the opinion of this author might constitute *the* genuinely Ultimate Limit of Physics [2]. As typical inhabitants of borderline territories, philosophers of all centuries (much before than physicists) have been squeezing their brains hoping to find a solution for this problem too which has been known among their community as the *mind vs. body* problem or *dualism vs. physicalism*. In this quest and especially in recent times many peculiar characters have born out of ingenious thought experiments across the philosophical literature thanks to the philosopher's vivid imagination. Famous among these characters are *Mary the color scientist* and the *philosopher's zombie* [3].

The present essay is an attempt to offer a quantum mechanical version of some of the most recently famous philosophical thought experiments related to the thorny debate of *mind vs. body* or *physicalism vs. dualism* (as you prefer…). Being led by Wigner's friend we will see that it is possible to study new interesting consequences arising from the "quantum versions" of characters like Mary and the philosopher's zombie. In the following sections a guided tour will be offered introducing all the *usual suspects* as well as their stories and if you continue reading even more, new situations will emerge by interchanging these actors. Perhaps (and this is the humble hope of the author) this *divertimento* could be interesting enough so to contribute somehow to the ongoing hunt for solving the mind-body problem.

Wigner's friend

Eugene Wigner is a very curious scientist that has a keen interest in measuring the state of quantum particles. He knows that quantum particles can only be measured in two states symbolized as $/+>$ and $/->$. Therefore his measuring device is constructed in such a way that produces a flash of light every time a quantum particle is detected in the state $/+>$ leaving the measuring device in a corresponding state symbolized by $/light+>$. On the other hand every time a quantum particle is detected in the state $/->$ it will leave the measuring device in a corresponding state symbolized by $/light->$ (no flash of light).

Let's suppose now that Wigner doesn't know what the state of the quantum particle that he wants to measure is initially. According to quantum mechanics Wigner will symbolize such a

situation as a *quantum superposition state* (i.e. *|particle>*) of the both alternatives where the particle can be:

$$\left|particle\right\rangle = C_+\left|+\right\rangle + C_-\left|-\right\rangle$$

Moreover if Wigner treats the combined system (particle + measuring device) as single quantum system, then he must express (following a rule of quantum mechanics known as linear superposition) the state of this combined system as:

$$\left|particle\right\rangle\left|device\right\rangle = C_+\left|+\right\rangle\left|light+\right\rangle + C_-\left|-\right\rangle\left|light-\right\rangle$$

Of course Wigner can check the outcome of a measurement that he performs to the particle at any moment by waiting to see whether his measuring device flashes or not. According to the quantum rules he will obtain a light flash with probability $\left|C_+\right|^2$ and no light flash with probability $\left|C_-\right|^2$. However, he chooses for a different approach and instead he steps out of the laboratory looking for a cup of coffee and requests from a friend to observe the result. After finishing his coffee, Wigner returns and asks his friend whether she saw a light flash (a YES state for his friend symbolized by *|YES>*) or not (NO state symbolized by *|NO>*). How should Wigner analyze situation *before his friend speaks*? Being an expert on quantum mechanics he considers his friend to be part of a larger system (particle + measuring device + friend) and applies again the linear superposition rule so the total system is represented by the superposition state *|particle>|device>|friend>*:

$$\left|particle\right\rangle\left|device\right\rangle\left|friend\right\rangle = C_+\left|+\right\rangle\left|light+\right\rangle\left|YES\right\rangle + C_-\left|-\right\rangle\left|light-\right\rangle\left|NO\right\rangle$$

Given this, Wigner can therefore anticipate that there will be a probability of $\left|C_+\right|^2$ that his friend will answer YES and probability of $\left|C_-\right|^2$ that she will answer NO. If his friend answers YES, then as far as Wigner is concerned, the state vector *|particle>|device>|friend>* **_collapses_** at that moment so the probability that a NO result was obtained is zero. Wigner therefore can infer that the particle was detected in the */+>* state and that the light of the device flashed. Not completely satisfied with this reasoning Wigner decides to go one step further on questioning his friend:

- **What did you feel about the flash <u>before I asked you</u>?** - Wigner says.

To which his friend replies:

- I did (did not) see the flash…don't bother me with such a stupid question…- his friend answers.

Since the answer is clear Wigner must conclude that his friend made up her mind about the outcome of the measurement **_before she was asked about it_**. Therefore the state vector that Wigner wrote *|particle>|device>|friend>* involving a superposition appears absurd to him since it will imply that his friend was in state of suspended animation before she answered his question in contradiction to what she is saying. On the other hand Wigner knows (because his expertise in quantum mechanics) that if his friend is replaced with a simple physical system (such as a single atom capable of absorbing light from a flash) then the mathematically correct

description must be in terms of the superposition and not either of the collapsed states. ***It follows, according to Wigner, that a being with consciousness must have a different role in quantum mechanics than an inanimate measuring device.*** Of course, there is nothing in principle to prevent Wigner from assuming that his friend was indeed in state of suspended animation before answering the question. However, to deny existence of the consciousness of his friend to this extent would be certainly an unnatural attitude to take. As far as Wigner knows he has a direct knowledge of his own sensations and while there is no strict logical reason to believe that others have similar experiences everybody admits that the phenomenon of sensations is widely shared by living organisms (Wigner calls this the *Principle of Charity*). Therefore he must believe his friend when she said that she had a determinate experience even before he asked her. *Wigner therefore must conclude that the superposition state collapses when interacting with the first conscious mind that it encounters*.

Mary the color scientist

The philosopher Frank Jackson devised an interesting thought experiment as an argument against physicalism (roughly speaking the view that the universe, including all that is mental, is entirely physical) [4]. The though experiment introduces a nowadays classic character called Mary who since her birth has led to many philosophical tangles. Mary is a specialist in the neurophysiology of color vision who lives in the far future when scientists already figured out all about the physical processes in the brain and how they produce behavior. In this sense she knows *everything there is to know* about color perception and the processing of color information in the visual system. Knowing everything there is to know will include of course the knowledge of how certain wavelengths of light stimulate the retina and produce the contraction of the vocal chords and expulsion of air that results in someone saying "this is a red apple". Nevertheless and for mysterious reasons Mary has been brought up all her life in a black and white room and forced to contemplate the world through a black and white television monitor never seeing any colors at all. One day Mary is let out of her black and white room by her captors and sees colors for the first time. We can imagine that her captors release her and immediately show her a red apple. At this moment Jackson proposes that one of the two alternatives happen:

Alternative 1: When liberated and after seeing the red apple Mary says: "Wow, I never realized red would look like *that*!"

Or on the contrary:

Alternative 2: When liberated and immediately after having seen the red apple Mary says: "OK, that's red, so nothing new for me of course…"

Accepting alternative 1 will mean that Mary learns something new that she didn't know before even though she was supposed to know everything there is to know about color perception. That something new will be her inner experience of *what red is like* (philosophers call it *qualia* or in singular *quale*) [3]. It will then imply that no amount of knowledge about the physical facts could have prepared her for the *raw feeling* of *what it is like* to see a red apple. In other words the physical facts about the world are not all there is to know. On the contrary, accepting alternative 2 will entail that the physical facts are indeed all there is to know about the world. The philosopher Daniel Dennett has championed this point of view arguing that the defenders of the first alternative simply fail to follow the instructions of the

story in allowing Mary to know *everything* there is to know about color. In this sense Dennett tells his own ending to Mary's story as [5]:

[...] And so, one day, Mary's captors decided it was time for her to see colors. As a trick, they prepared a bright blue banana to present as her first color experience ever. Mary took one look at it and said 'Hey! You tried to trick me! Bananas are yellow, but this one is blue!" Her captors were dumfounded. How did she do it? "Simple," she replied. "You have to remember that I know everything–absolutely everything–that could ever be known about the physical causes and effects of color vision. So of course before you brought the banana in, I had already written down, in exquisite detail, exactly what physical impression a yellow object or a blue object (or a green object, etc.) would make on my nervous system. So I already knew exactly what thoughts I would have (because, after all, the 'mere disposition' to think about this or that is not one of your famous qualia, is it?). I was not in the slightest surprised by my experience of blue (what surprised me was that you would try such a second-rate trick on me).

The debate about Mary vividly continues and has generated a wide range of alternatives and points of view impossible to account here. Nevertheless the bottom line is that if you believe that Mary will learn anything new when she comes out, then you believe that consciousness, subjective experience, or qualia, are something additional to the knowledge of the physical world. On the other hand if you think she will not learn anything new, then you believe that knowing all the physical facts tells you everything there is to know, including the subjective experience something. Briefly stated and roughly speaking if you think Mary will be surprised you are compelled to reject physicalism and adopt dualism.

Evil Wigner and quantum Mary

So far we have introduced two peculiar characters: Wigner's friend and Mary the color scientist. Both of them have helped since their appearance to illustrate and stimulate the extensive and controversial debate over the nature of consciousness and its role. Nevertheless both characters were conceived as playing a fundamental role in separate stories. Our intention here is to study the consequences of mixing them up in a new though experiment that will incorporate as we might say "the best of the two worlds". In doing that let's consider a new alternative involving the quantum physicist Wigner and his new quantum girlfriend Mary:

After experimenting with his friend, Wigner became a sort of evil scientist and decided to continue with Mary since he kept her locked in a black and white room for all her life. Similar to the quantum particle test but more maliciously, this time in the room Wigner places a radioactive element that decays or doesn't decay after a certain amount of time (states $/+>$ and $/->$ respectively). Coupled to the radioactive element there is a device provided with two doors. If the radioactive element decays then door number 1 opens showing Mary a red apple. If the radioactive element doesn't decay door number 2 opens showing Mary a blue banana (states $/R>$ and $|B>$ respectively). According to Dennett's argument only one of the following two alternatives is possible:

Alternative 1: Mary, after seeing the red apple, won't show any surprise (state $/NS>$) because *she knows all there is to know* about color experiences.

Alternative 2: Mary, after seeing the blue banana, will show surprise (for trying to be fooled by her captor Mr. Wigner and symbolized by state /S>) because again *she knows everything there is to know* about color experiences.

Just as in the case of his friend, Wigner considers the complete superposition state /element>/device>/Mary> *before speaking to Mary about her perceptions*:

$$\left|Element\right\rangle\left|Device\right\rangle\left|Mary\right\rangle = C_{+}\left|+\right\rangle\left|R\right\rangle\left|NS\right\rangle + C_{-}\left|-\right\rangle\left|B\right\rangle\left|S\right\rangle$$

And again Wigner asks Mary:

- **What did you feel <u>before I asked you</u>?**

To which she replies:

- I did feel surprised because you were trying to fool me with the blue banana trick and I know everything there is to know about color experiences (alternatively, I didn't feel surprised since all that I saw was a red apple and I know everything there is to know about color experiences).

Like in the case of his friend, Wigner is entailed to conclude that unlike he wants to deny the existence of Mary's consciousness he should admit that her consciousness was causing the collapse of the wave function of the superposition state. Even more and strangely enough a lurking dualism have re-appeared even though we tried to carefully putting it away by allowing Mary to know *everything there is to know* about color perception as the philosopher Daniel Dennett suggested. A simple conclusion will be to state that Dennett's argument is failing at some point and after all we are therefore impelled to accept dualism. The author would like to leave the debate open at this point but not without considering that perhaps what the quantum version of Mary points to as a plausible alternative would be to catalog the though experiment of *Mary the color scientist* as an inadequate tool for elucidating the question of dualism vs. physicalism.

Wigner's zombie detector

Enough about Mary … let's now consider another of the characters that certainly has caused more troubles than her: The philosopher's zombie (*p-zombie* from now on). The p-zombie is someone who looks like you, acts like you, speaks like you, and in every detectable way behaves exactly like you, but is not conscious [6] . In other words a p-zombie is a hypothetical being that is except that it lacks conscious experience; if asked the p-zombie will not hesitate to answer "of course, I am conscious". The question whether p-zombies are conceivable or not has generated a tremendous amount of debate that in some cases turned out to be complicated to follow [7]. In simple terms the whole debate could be stated like this: On the one hand, if you believe in the possibility of conceiving p-zombies then you believe that consciousness is some kind of inessential optional extra to behavior. On the other hand, if you believe that p-zombies are not possible to conceive you must believe that anything that could perform all the behaviors we perform would necessarily be conscious.

Eugene Wigner, our controversial hero scientist, is not taking any sides in the conceivability debate but he thinks that he has invented a p-zombie detector which according to him will demonstrate that p-zombies are definitely distinguishable from normal human beings. As in

the case of his friend and using his measuring device this time Wigner decides to ask a hypothetical p-zombie *(hp-zombie)* to observe the result of an experiment for detecting the state of a quantum particle while leaving the room. Accordingly, Wigner will treat the superposition state */particle>/device>/hp-zombie> before speaking to the hp-zombie about his perception as*:

$$\left| particle \right\rangle \left| device \right\rangle \left| hp - zombie \right\rangle = C_{+}\left| + \right\rangle \left| light + \right\rangle \left| YES \right\rangle + C_{-}\left| - \right\rangle \left| light - \right\rangle \left| NO \right\rangle$$

After entering the room again Wigner asks the hp-zombie as usual:

- **What did you feel about the flash <u>before I asked you</u>?**

This time if the hp-zombie turns out to be a true p-zombie (i.e. with a lack of consciousness) and since Wigner conclusion after previous experiments is that consciousness is the very one responsible for the collapse of the state vector the answer will be something like:

- I am afraid I haven't perceived yet a definite outcome of the experiment…

It turns out that Wigner's device apart of being able to measure the state of a quantum particle, is genuinely a fantastic p-zombie detector. In such a way we have, thanks to Wigner's device, a perfectly operational way to distinguish p-zombies from normal human beings possessing consciousness. Are therefore p-zombies conceivable? Maybe…are they undistinguishable from you and me? If we follow Wigner's argument, definitively not. Again the author would like to leave the debate here not without questioning again the suitability of the p-zombie construction in elucidating the controversy of mind vs. body.

<u>Epilogue</u>

In the past sections we have offered a quantum mechanical description of some of the most remarkable thought experiments arising in the recent philosophical literature about the debate of mind vs. body. Our starting point has been the well known quantum story of Wigner's friend and we have fabricated other examples based on it. The reason why the author has chosen as starting point Wigner's argument and conclusions is his belief that none of the present alternative theories concerning the so called measurement problem in quantum mechanics has been able to offer an definite explanation of it (including the notable efforts based on decoherence) [8]. Therefore Wigner's thesis seems to be as valid as ever. Not all the consequences of the new proposed quantum versions of the classical philosophical though experiments have been considered here. Nevertheless some of the outlined conclusions point to question the validity of the classical philosophical thought experiments as useful tools in the controversy of mind vs. body.

Finally, along these lines the fundamental question to be elucidated according to the author can be simply stated like this:

Is the following equation true?

Quantum Mechanics + Principle of Charity = Dualism

Physics (especially quantum mechanics) as a theory of how and what is possible to be known should be able to illuminate this ultimate question.

References

[1] Wigner, E.P. (1961) "Remarks on the mind-body question." In The Scientist Speculates, ed. I.J. Good, 284-302. (Heinemann).

[2] H. Atmanspacher (2004). Quantum Theory and Consciousness: An Overview with Selected Examples. *Discrete Dynamics in Nature and Society* 1 (2004) 51-73.

[3] Blackmore, S. (2003) *Consciousness: An Introduction*, Oxford University Press.

[4] Jackson F. "Epiphenomenal Qualia" *Philosophical Quarterly*, 32 (1982) 127-36.

[5] Dennett, D.C. (1991), *Consciousness Explained,* Boston, MA: Little, Brown & Co.

[6] Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press.

[7] Kirk, Robert, "Zombies", *The Stanford Encyclopedia of Philosophy (Summer 2009 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2009/entries/zombies/>.

[8] Barrett, J.A. (1999). *The Quantum Mechanics of Minds and Worlds*, Oxford University Press.