Article

Sleeping Beauty in Quantumland

Na'ama Hallakoun*, Berry Groisman[†,‡] and Lev Vaidman*

ABSTRACT

The Sleeping Beauty Problem is analysed within the framework of the Many-Worlds Interpretation of Quantum Theory. The view of the Many-Worlds Interpretation as fundamentally deterministic theory with no place for objective uncertainty is advocated. The main concepts of this version of the Many-Worlds Interpretation are reviewed and it is shown how their application leads to a *Thirder* solution to the Sleeping Beauty Problem, and not to the *Halfer* solution as it was claimed by Peter Lewis.

Key Words: decision theory, subjective uncertainty, many worlds interpretation, quantum sleeping beauty

NeuroQuantology 2013; 1: 149-159

1. Introduction

The Many-Worlds Interpretation of Quantum Mechanics (MWI) is a natural, yet relatively unexplored, playground for the Rational Decision Theory. In the MWI the nature puts all the cards on the table, which allows one to treat all possible scenarios faced by rational agents on equal footing. One of the first attempts in this direction (Vaidman, 2001) was made in relation to the *Sleeping Beauty Problem (SBP)*, which shares a similar challenge with the MWI: both are expected to provide a consistent account of an agent's subjective uncertainty on the background of an underlying objective certainty.

There is no consensus on the SBP amongst (classical) decision theorists. In the past we have analysed the problem in the classical case (Vaidman and Saunders, 2001; Groisman, 2008). In this article we analyse the problem within the MWI framework, claiming for the same answer as in the classical case, as was first proposed in (Vaidman, 2001). In our view, even though the classical setup might not offer a clear solution accepted by everyone, the MWI provides a much more compelling argument. We compare our treatment of the problem with the one presented by Peter Lewis, who advocates a different solution (Lewis, 2007; 2009).

The main rationale for analysing problems like SBP within the MWI framework lies in the difficulties with conceptual interpretation of classical probability. We lack solid foundations for classical probability theory. The measure theory, on the other hand, provides a measure on a set with clear conceptual status. The MWI framework allows us to give up classical probability in favour of a measure.

This paper is organised as follows: Section 2 provides the reader with a brief review of the Sleeping Beauty Problem. Section 3 introduces the Many-Worlds Interpretation of Quantum Theory and discusses the problem of probability within its framework. Section 4 addressed the question of the treatment of SBP within the framework of the MWI. Finally, Section 5 summarises the results of the paper and points out direction of further research.

Corresponding author: Lev Vaidman
Address: *School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel, naamah1@post.tau.ac.il, vaidman@post.tau.ac.il. [†]DAMTP, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK. [‡]Sidney Sussex College, Sidney Street, Cambridge CB2 3HU, UK, b.groisman@damtp.cam.ac.uk
vaidman@post.tau.ac.il
Received Feb 1, 2013. Accepted March 13, 2013.
eISSN 1303-5150

2. The Sleeping Beauty Problem

In 2000, Adam Elga (2000) published a paper concerning a rather philosophical problem based on an unpublished work by Arnold Zuboff. The problem, known as the "Sleeping Beauty Problem" (SBP), can be stated as follows:

> Sleeping Beauty is put to sleep on Sunday. Beforehand, she is told that she will be woken either once or twice, depending on the result of a fair coin toss: If Heads – only once on Monday, if Tails – once on Monday and once on Tuesday. In the end of the first awakening she is given an amnesia drug, which erases her memory of this awakening, and put back to sleep. Upon each awakening, Beauty is asked to what degree she believes the outcome of the coin toss was Heads.

The problem looks quite innocent and straightforward. Elga claimed her answer should be "1/3". His argument went roughly as follows: If the experiment is repeated many times, in 1/3 of the awakenings the coin will land Heads. Thus, Beauty's credence in the coin landing Heads in *any* particular awakening, should be 1/3 (Elga, 2000). This argument is based on the frequentist definition of probability. Elga also provided an elaborate Bayesian account of his position, which will be presented below.

Most likely this would have settled the issue, if the prominent philosopher David Lewis had not contested Elga's argument (Lewis, 2001). In a nutshell, Lewis's argument goes as follows: The probability of getting a Heads result when tossing a fair coin is 1/2. Since Beauty knows from the beginning that she is going to wake up at least once, and since she cannot recall her previous awakenings (if any), she gains no new knowledge during the experiment. Thus, her credence in the coin landing Heads should remain 1/2.

This sparked quite a controversy amongst the researchers in rational Bayesian decision theory. The opponents split into two rivalling groups: the majority thinks Beauty should answer "1/3", hence they are called "*Thirders*", and the minority thinks she should answer "1/2", hence – "*Halfers*".

Let us turn now to a more detailed account of the disagreement between Lewis and Elga:

According to Elga (2000), when Beauty wakes up, she finds herself in one of the following three possibilities:

- H_1 – Heads and it is Monday.
- T_1 – Tails and it is Monday.
- T_2 – Tails and it is Tuesday.

We define a "*centred proposition*" as a proposition which refers to a specific individual in space and/or time. Thus, while both T_1 and T_2 describe the same *uncentred* situation (i.e., the coin has landed Tails), they correspond to two different *centred* propositions.

Elga argued that upon awakening Beauty's credence in Heads should change from initial $P_S(H)$=1/2 (the S is for "Sunday") to $P(H)$=1/3. He justifies this change by Beauty's transition from a situation in which she counts her temporal location as irrelevant to the truth of Heads, to one in which she counts her temporal location as relevant to the truth of Heads. Since the three possible awakenings are indistinguishable, he deduced that they have equal probabilities $P(H_1)=P(T_1)=P(T_2)$ and associates $P(H)$ with $P(H_1)$, so $P(H)$=1/3.

In his reply David Lewis (2001) criticised Elga's Thirder approach. According to Lewis, Beauty receives no new information, centred or uncentred, during her awakening, because she already knew she will be awakened, and she still does not know which one of the three possible awakenings she is in. Hence, her credence in Heads should remain ½.

3. Probability in the Many-Worlds Interpretation

One of the central conceptual difficulties faced by Quantum Mechanics is the Measurement Problem, which constitutes in defining a boundary between two processes. One is a process of deterministic evolution of a quantum wavefunction between measurements, which is governed by the Schrödinger equation. The other is the evolution of the wavefunction during measurements, which appears to be non-deterministic. Different ways to approach this problem have been developed, which resulted in several rival theories within Quantum Mechanics (known traditionally as Interpretations). The first and the most traditional among these interpretations is the Copenhagen Interpretation, established and actively advocated by Niels Bohr. According to this interpretation, there is a genuine *uncertainty* regarding results of quantum experiments to be conducted. When a measurement is performed on a quantum system only one of the possible outcomes is actualised. This phenomenon is known as the *collapse* of a quantum wavefunction. The system's wavefunction, before the measurement, can be written as a linear superposition of observable's eigenstates, each with a certain amplitude. The probability of obtaining one of the possible results equals the squared absolute value of the corresponding amplitude. This constitutes the *Born Rule*. The Universe described by the Copenhagen Interpretation is both non-deterministic and non-local – both features perceived as a major deficiency by some physicists, including Albert Einstein.

One of the central theories rivalling the Copenhagen Interpretation is the Many-Worlds Interpretation (MWI), first suggested by Everett in 1957. In broad terms the idea of the MWI is to include the experimentalist-observer (and the entire Universe) in the description of the measurement process, such that the observer, together with the measurement device, enters into a correlated (entangled) state with the measured system. The entire system-device-observer combination keeps evolving according to the Schrödinger equation with the observer's mental state corresponding to seeing different outcomes in different branches of the superposition. There is no place for collapse in the MWI: all the possible outcomes are realised, and the world splits into a number of actual worlds after each measurement performed. Thus, the quantum theory, through the MWI, recovers both determinism and locality (Vaidman, 2002).

The proponents of the MWI have encountered significant challenges, which resulted in several versions of the theory being developed. One of the major issues is the treatment of probability, on which there is no consensus amongst supporters of different versions.

3.1 The Universe Without Genuine Probability

According to one view, which has been developed by one of us (Vaidman, 1998) and which we advocate here, there is no genuine probability in the MWI framework as it is a completely deterministic theory and there is no relevant information that an observer who is going to perform a quantum experiment is ignorant about. The quantum state of the Universe at one time specifies the quantum state at all times. Nevertheless, we have to reconcile this with our everyday perception of probability. This can be achieved by associating an "*illusion of probability*" with post-measurement *ignorance* (Vaidman, 2012). In the *Sleeping Pill experiment* (Vaidman, 1998) the agent is given a sleeping pill and sleeps through a quantum coin toss, e.g. an experiment with a spin-half particle prepared in an equal superposition of eigenstates of a measured operator. While asleep she is moved to either room A or an identical-looking room B, based on the result of the toss. It is meaningless to ask her *before* the toss what is the probability to find herself in room A after the toss. Since she is the ancestor of both possible future descendants who will exist simultaneously, she knows it will happen with certainty. However, we *can* ask one of the descendants after the toss, but *before* she is aware of its outcome, in which room does she think she is in. The illusion of probability follows from the identity between the observer's memory state in a hypothetical Collapse scenario and her memory state in the MWI scenario. In other words, an observer in a hypothetical Collapse-Universe, repeating a certain experiment several times, will have the same memories as an observer performing similar experiment in a MWI-Universe (Vaidman, 2012). In this interpretation, the squared absolute value of the amplitude of a world is called its "*measure of existence*" (henceforth referred to as *the measure*). In the context of future worlds the measure acts as the "*caring measure*", which reflects how much one should "care" about his future (post-split) descendants. The "*behaviour principle*" teaches us that one should care about his descendants according to their measures, and thus functions as an equivalent to the Born Rule in the MWI. Paul Tappenden (2011) has recently named this the *Born-Vaidman Rule*.

In the Sleeping Pill experiment described above, before the coin toss, the agent is certain that there will be an agent psychologically continuous with her, who will be placed in room *A*, but will be genuinely uncertain about her location. Likewise, there will be an agent psychologically continuous with her, who will be placed in room *B*, but will be genuinely uncertain about her location. Thus, each of the agents is located in one of the worlds (branches) and experiencing only one of the two outcomes of the coin toss. The fact that there is no direct meaning to the probability of the outcomes of the experiment does not contradict a genuine uncertainty of the two successors upon awakening in propositions "I am in *A*-world" and "I am in *B*-world".

The above treatment of probability in the MWI is far from being widely accepted. Alternative approaches are centred around assigning pre-branching uncertainty. In particular, Saunders and Wallace (Saunders, 1998; Wallace, 2006; Saunders and Wallace, 2008) introduced branching structure which makes the observer genuinely uncertain about what will happen to her. We turn now to their ideas (for a critical discussion of Subjective Uncertainty see Tappenden, 2008; Kent, 2010).

3.2 I'm Uncertain, Therefore I Am?

3.2.1 Saunders-Wallace's Argument for Subjective Uncertainty

The concept of *Subjective Uncertainty* (or *Subjective Indeterminism*) was first suggested by Simon Saunders (1998), and later expanded by David Wallace (2006). According to the Saunders-Wallace argument, there is a genuine uncertainty in the MWI, and an observer can assign genuine pre-branching probabilities to the outcomes of an experiment. They base their claim upon the following argument:

Consider an observer who is about to perform a quantum experiment with two (or more) possible outcomes. What should this observer expect before the experiment? There are two options we shall consider:

1. She should expect to become both future selves.

2. She should expect to become either one future self (but not the other).

According to Saunders and Wallace option (2) is the only viable one and, as Wallace puts it (Wallace, 2006), "In the absence of some strong criterion as to which copy to regard as *'really'* me, I will have to treat the question of *which* future self I become as (subjectively) indeterministic."

Option (1) is rejected on the grounds that it requires some impossible telepathic link between the future selves of the observer. Wallace argues that such link "will have to supervene on some physical interaction between the two copies – an interaction which is not in fact present". For a discussion of the possibility of an interaction between future selves see (Mensky, 2007). We agree with Saunders and Wallace that such a telepathic link between the descendants of the observer is implausible – the observer is indeed not a Hydra-like creature (The Hydra metaphor follows Tappenden's example (Tappenden, 2011)), sending and controlling its many heads all over the worlds. But no such interaction is needed! The relation between each of the future selves and the pre-branching observer is identical: They both have the same history (the history of the pre-branching observer), and prior for realising in which world they are in, they also have the same state of mind (and an illusion of probability regarding the answer).

In the light of the above conclusion, we believe option (2) is illegitimate: what an agent identifies as future "I" is not a legitimate question in the MWI. The future "I"s embrace the entire wavefunction of the Universe and one can only trace the "trajectory" of "I" through branching history backwards in time.

As described by Saunders and Wallace (Saunders and Wallace, 2008; Saunders, 2010), the worlds are *diverging* rather than *overlapping* (branching): While overlapping worlds picture the process of *branching*, in which one world splits into several others, diverging worlds picture *separate* worlds, which exist prior to the measurement, but are merely indistinguishable before the measurement takes place. The diverging worlds concept introduces a true *ignorance* probability into the theory, since it implies that each possible world exists prior to the measurement, meaning that although indistinguishable, there are as many copies of the observer before the measurement process, as there are after it. Each copy of the observer is truly ignorant, *before* the experiment takes place, as to which world she belongs. In our overlapping (branching) picture, however, before the measurement there is only one copy of the observer who is about the perform the measurement. Since she has nothing to be ignorant about before performing the measurement, there is no genuine uncertainty, but only an *illusion* of probability.

3.2.2 The Many-Minds Interpretation

Saunders and Wallace's Subjective Uncertainty brings to mind another variation of the MWI, namely – the *Many-Minds Interpretation* (MMI). According to the MMI, as suggested by David Albert and Barry Loewer in 1988 (Albert and Loewer, 1988), there is an infinite pool of possible Minds (non-quantum mechanical objects related to our mental state), that exist prior to the measurement. Before the experiment (say, measuring the *x*-component of an electron's spin), the brain of the observer is in a state in which no Mind has a certain belief about the result of the upcoming experiment. After the measurement process, the combined quantum state of the observer's Brain and the electron is:

$$c_1 |B_{\uparrow_x}\rangle \otimes |\uparrow_x\rangle + c_2 |B_{\downarrow_x}\rangle \otimes |\downarrow_x\rangle. \qquad (1)$$

Albert and Loewer postulate that the coefficients c_i follow Born Rule, so the proportion of Minds observing spin-up is $|c_1|^2$, and the proportion of Minds observing spin-down is $|c_2|^2$. Since this was known to each of the Minds of the observer before the measurement process, each Mind has credence of $|c_1|^2$ in observing spin up. Thus, in order to settle the probability problem, Albert and Loewer introduce an additional postulate, which leads to genuine probability. In fact, the MMI can be described as a *hidden variables* theory, in which the hidden variables are the Minds, as can be learned from the following quotation (Albert and Loewer, 1988):

> "An infinity of minds is required since a measurement or a sequence of measurements may have an infinite number of outcomes. Furthermore, although the evolution of individual minds is probabilistic, the evolution of the set of minds associated with *B* [the state of the Brain] is deterministic since the evolution of the measurement process is deterministic and we can read off from the final state the proportions of the minds in various mental states."

Reflecting back on the Saunders-Wallace argument, we can see that it shares similarity with the MMI: according to Saunders and Wallace the observer should expect to become either one of her future selves, "and in the absence of some strong criterion as to which copy to regard as "really" [her], [she] will have to treat the question of *which* future self [she] become[s] as (subjectively) indeterministic" (Wallace, 2006). Such observer is pretty much similar to Albert and Loewer's Mind: the measurement process itself is deterministic, but it is assumed that there is an indeterministic process from the viewpoint of the observer's Mind. Moreover, the divergent picture described in Section 3.2.1 is similar to the MMI's continuum of Minds. Wilson (2012) develops the idea of "diverging worlds" by introducing "indexicalism". He distinguishes between worlds by their index and argues that the MWI involves a set of distinct "parallel" worlds with each observer located in one world only. Since quantum mechanics does not provide any physical candidate for such an index, it seems that all these attempts are equivalent to the proposal of Albert and Loewer with their non-physical mind serving as an index. Thus, we find a very close similarity between the Subjective Uncertainty view and the MMI.

3.2.3 On Determinism and Rationality

Let us recall that we do not find it difficult to reconcile between an observer's certainty that all the possible outcomes of a quantum experiment will be realised and the subjective uncertainty of her descendants after the experiment as to which world in the Universe they are in. As a matter of fact, this apparent difficulty is one of the reasons for the disagreement on the MWI-version of the SBP, as we claim below, in Section 4.1.

It is worth pointing out that the question the observer asks herself in the Saunders-Wallace argument ("What should I expect to *become*?") is not well-defined in the framework of the MWI. The observer can ask herself two different questions prior to the measurement: "What should I expect that will *happen*?", and "What should I expect to *feel*?". The answer to the question "What should I expect that will *happen*?" in the framework of the MWI is clear: since the MWI is completely deterministic, she should expect that all possible outcomes are realised, thus rendering this question pointless. The answer to the question "What should I expect to *feel*?" is clear as well: since she has no interaction with her counterparts (as agreed by Saunders and Wallace), she should expect to experience only one single outcome, in each world. This was also pointed out by Hilary Greaves (2004): "Not that [she] should expect to see *both*: [she] should expect to see *each*."

However, since the observer knows that the Universe is deterministic, and that every possible option is realised, one might fear she starts to act irrationally. Peter Lewis (2000) argued that a MWI advocate would agree to play a *quantum Russian roulette* game, no matter how many times, since there will always be a world in which she lives on, happily ever after. This is where the *behaviour principle* comes in. The goal of the behaviour principle is to ensure a rational behaviour of advocates of the MWI. In fact, one can say it is a rather phenomenological concept we use to explain people's behaviour. But this is not unique to the MWI: We use a similar rule – the Born Rule – in the framework of Collapse interpretations, in which observers are expected to place bets according to the amplitudes of the possible outcomes. Thus, the observer should avoid participating in the game, since the measure of the worlds in which she is dead is much greater than that of the worlds in which she is alive (Vaidman, 2002).

Is the behaviour principle only a phenomenological feature? Does the measure of her world have an influence on the observer? As was demonstrated by one of us (Vaidman, 1998), one can think of a gedanken experiment with an interference of *macroscopic* objects. Such an experiment requires super-technology, which currently we do not possess. However, we can perform an interference experiment with microscopic particles: For example, we can take a neutron and place it in a Mach-Zehnder interferometer. The interferometer splits the neutron's world into two, and then reunites them. As it is shown in (Vaidman, 1998), the measure of each of the neutron's copies influences the outcome of the experiment. This microscopic experiment is analogous to the gedanken experiment of macroscopic objects' interference.

But how does all this connect to the debate about Subjective Uncertainty? The benefit of introducing pre-branching assignment of probability constitutes in the possibility to use traditional decision theory arguments in the face of branching without modifications (Wallace, 2006). We argue, however, that an advocate of the MWI, facing branching, should act as rationally as anyone else, even though she has nothing to be uncertain about. All that is needed is the behaviour principle, which defines the meaning of the Born Rule within the framework of the MWI. Our claim is strengthened by Tappenden, who argues that "The very possibility of applying the Born–Vaidman rule post-branching to assign credences to outcomes can serve as a substantive guide to pre-branching action" (Tappenden, 2011).

With this in mind, we can finally turn to discuss the SBP within the framework of the MWI.

4. Sleeping Beauty in the Many-Worlds Universe

Elga's and Lewis' papers have initiated a long-lasting debate regarding the (classical) SBP and to the best of our knowledge no consensus has been reached yet. It was suggested by one of us (Vaidman, 2001) to address the problem in the framework of the MWI.[2] The treatment of the SBP within the MWI generated almost as much controversy as the original Elga's scheme, with no consensus on the credence (Lewis, 2007; Papineau and Durà-Vilà, 2009a; Lewis, 2009; Papineau and Durà-Vilà, 2009b; Peterson, 2011; Bradley, 2011; Wilson, 2013).

4.1 Treatment of the SBP in the Framework With No Genuine Probability

The SBP becomes *Quantum* SBP when the fair coin is replaced by a fair quantum coin. There is a clear benefit in using a quantum coin – it is an ultimate fair coin. We believe, nonetheless, that the nature of the coin does not affect the solution to the problem. This assumption might give an impression that after the classical coin is replaced by a quantum one the problem (or the absence of it) stands as it is, since the technological realisation of the coin toss is not crucial.

[2] This idea had arisen during discussions with Simon Saunders.

However, there are important advantages in using a quantum coin and analysing the problem in the MWI framework.

The apparent contradiction between the reasoning of *Halfers* and *Thirders* in the standard classical SBP rests on the failure to distinguish two different questions posed to Beauty. On Sunday Beauty can be asked two distinct questions: One is "what is your credence in the future result of the coin toss being Heads?". The other is "what will be your credence in H_1 upon awakening?" or, to rephrase, "what will be your degree of belief in Heads upon awakening in the world you are asked the question in?". In the MWI this distinction emerges naturally as a matter of necessity, because the question "what is your credence in the future result of the coin toss being Heads?" is meaningless – both options are realised. According to the version of the MWI advocated by us, there is no direct meaning to the probability before the experiment: all the information is available to the observer before the measurement. However, the descendants *are* ignorant about the outcome (Vaidman, 2001). Thus, the second question, "what will be your degree of belief in Heads upon awakening in the world you are asked the question in?", makes perfect sense (see Section 3). It corresponds to the *centred* question "What is Beauty's credence in Heads in the awakening she's asked the question in?". This has far reaching ontological implications. Within the framework of the MWI one loses the feature of the classical SB experiment, that David Lewis found so objectionable, namely that Beauty changes her credence in an uncentred proposition, although she did not receive any uncentred evidence: In the MWI all propositions relevant to the SBP scenario are centred propositions.

We now turn to the explicit analysis of the problem. First, let us start with a modified experiment. Let us assume that Beauty is awaken twice regardless of the result of the coin toss (with a memory erasure pill given to her after the first awakening, which causes her to forget the Monday awakening completely). Beauty knows that in the Universe there are four different events – the four possible awakenings (H_1, H_2, T_1, T_2), as shown in Figure 1(a).

Let us write explicitly the corresponding wavefunction of the Universe at two times t (Monday) and t' (Tuesday) as follows

$$|\Psi_U(t)\rangle = \frac{1}{\sqrt{2}}|\mathrm{T_1: SB\ awaken}\rangle + \frac{1}{\sqrt{2}}|\mathrm{H_1: SB\ awaken}\rangle,$$
$$|\Psi_U(t')\rangle = \frac{1}{\sqrt{2}}|\mathrm{T_2: SB\ awaken}\rangle + \frac{1}{\sqrt{2}}|\mathrm{H_2: SB\ awaken}\rangle. \quad (2)$$

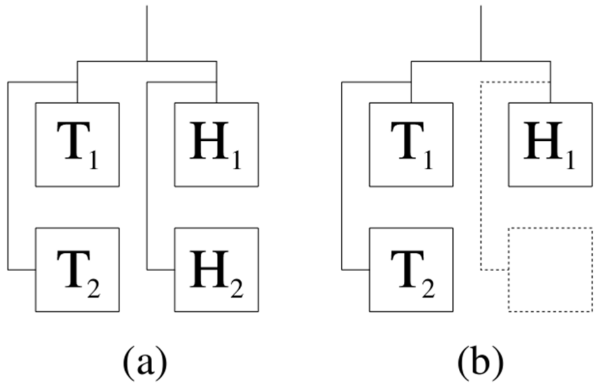

Figure 1. (a) Sleeping Beauty Experiment with four awakenings. (b) The original scheme with three awakenings. The lines in (a) and (b) indicate a psychological continuous connection.

All four measures corresponding to the four possibilities (Beauty's subjective worlds) are equal because the moduli squared of the amplitudes of the corresponding terms in the wavefunction of the Universe at times t and t' are equal. As Beauty's credences in each possibility should sum up to one, she should assign 1/4 to each, that is 1/2 that she is in the Heads-branch. This conclusion is uncontroversial and it is not contested by the classical rational decision theorists.

Now we turn to the original scheme. The only difference from the scheme described above, is that H_2 is removed, as shown in Figure 1(b). The presence or absence of H_2 obviously cannot affect the arguments for the equality of the measures of H_1 and T_1 and, similarly, T_1 and T_2. Indeed, this is reflected in the following change in the expressions for the wavefunction

$$|\Psi_U(t)\rangle = \frac{1}{\sqrt{2}}|\mathrm{T_1: SB\ awaken}\rangle + \frac{1}{\sqrt{2}}|\mathrm{H_1: SB\ awaken}\rangle,$$
$$|\Psi_U(t')\rangle = \frac{1}{\sqrt{2}}|\mathrm{T_2: SB\ awaken}\rangle + \frac{1}{\sqrt{2}}|\mathrm{H_2: SB\ asleep}\rangle. \quad (3)$$

The amplitudes of the corresponding branches are equal. Thus, the measures of her *subjective* worlds in all three events are equal.

Since only one of them corresponds to Heads she should answer "1/3" (Vaidman, 2001; 2012). This is consistent with Elga's conclusion – as Beauty has no reason to assign different (subjective) measures to any of the three possible awakenings (H_1, T_1, T_2), she should assign equal credence to each (Elga, 2000).

Let us now turn to the discussion of ideas of Peter Lewis (2007), who argues that the treatment of the problem with a quantum coin toss within the MWI leads to the answer 1/2.

4.2 Quantum Sleeping Beauty According to Peter Lewis

Lewis proposed a modified version of the experiment with *two* quantum coin tosses. If the first coin lands "Heads" then Beauty is awaken on Monday only. If it lands "Tails", then the second coin is tossed. Subsequently, Beauty is awaken on either Monday or Tuesday depending on the respective outcomes, "Heads" or "Tails", of the second toss. Thus, each combination of coin tosses results leads to only one awakening of Beauty.

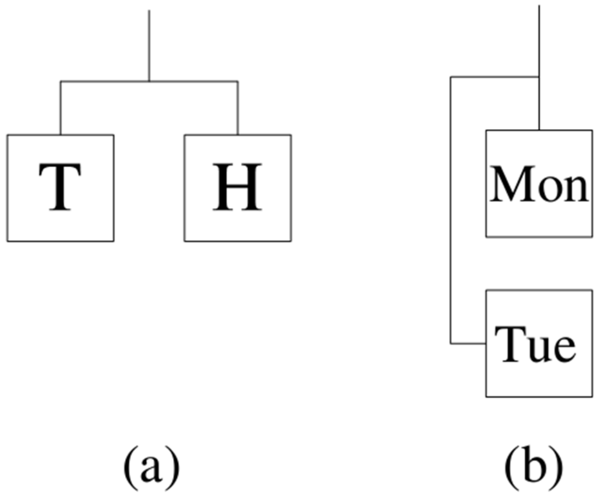

Figure 2. (a) Simple (fair) quantum coin toss experiment; (b) Simplified SB experiment: Beauty's subjective experience resembles the one in (a). The lines in (a) and (b) indicate a psychological continuous connection.

First, Lewis considers a simplified version of the (classical) SB experiment, in which there is no coin toss at all, and Beauty is awaken once on Monday and once on Tuesday, while not remembering the Monday awakening. (The simplified version corresponds to the chain of events in the Tails-world of the original experiment, after the coin toss.) Upon awakening, Beauty is asked for her credence that today is Monday. Lewis observes that there are "strong structural parallels between the Everettian treatment of a simple spin measurement and the simplified version of the Sleeping Beauty puzzle" (Lewis, 2009). Based on this analogy, Lewis then argues that "the same subjective probability assignment is appropriate in each case to quantify the agent's uncertainty", which motivates him to treat the original (single-coin) SBP setup and the two-coin scenario as equivalent in terms of the values of corresponding credences.

The "illusion of probability" concept, especially in relation to the "Sleeping Pill" experiment (Vaidman, 1998) mentioned in Section 3.1, has a striking resemblance to the simplified SBP. Because of the memory erasure pill, Beauty's epistemic situation is identical to that of the observer in the Sleeping Pill experiment. Therefore, we agree with Lewis that in the simplified classical version Beauty ought to have the same "post-branching" degree of belief in the propositions "This is a Monday/Tuesday awakening" as in corresponding propositions "This is Heads-branch/Tails-branch" in the simple quantum coin (or Sleeping Pill) experiment (see Figure 2). The measures of the two branches in the quantum coin experiment are equal. Similarly, in the simplified SB the two awakenings correspond to two subjective worlds in a single world at two different times and their measures are equal. It is important to make two following remarks. First, note that the Sleeping Pill quantum coin is similar to awakenings with memory erasing pill (simplified SB) only when the quantum coin is not biased. Second, this similarity follows from identical "post-branching" degrees of belief in the two settings, which in turn follows from equality of the measures within each setting. There is an important difference, however. In the case of simplified SB the measures equal one and have to be normalised giving the subjective degrees of belief of 1/2, while in Sleeping Pill quantum coin setup the measures themselves equal 1/2. This difference explains why the two setups cannot be treated as equivalent when used as a part of a larger experiment (see below).

We further agree with Lewis that Beauty's credence in Heads in the two-coin version of the experiment ought to be 1/2. This is supported by the analysis based on the measure. With two coins Beauty knows that in the Universe there are three different events – the three possible awakenings, namely A, B and C corresponding to $H^{(1)}$, $T^{(1)}H^{(2)}$ and $T^{(1)}T^{(2)}$ respectively, as shown in Figure 3 (here $H^{(i)}$ corresponds to Heads of the first coin and so on). As a result, after the two quantum coin tosses, the state of the Universe becomes

$$|\Psi_U\rangle = \frac{1}{\sqrt{2}}|\text{A}-\text{branch}\rangle + \frac{1}{2}|\text{B}-\text{branch}\rangle + \frac{1}{2}|\text{C}-\text{branch}\rangle, \quad (4)$$

which can be written more explicitly at two times, t (Monday) and t' (Tuesday) as

$$|\Psi_U(t)\rangle = \frac{1}{\sqrt{2}}|\text{A: SB awaken}\rangle$$

$$+\frac{1}{2}|\text{B: SB awaken}\rangle + \frac{1}{2}|\text{C: SB sleeps}\rangle, \quad (5)$$

and

$$|\Psi_U(t')\rangle = \frac{1}{\sqrt{2}}|\text{A: SB sleeps}\rangle$$

$$+\frac{1}{2}|\text{B: SB sleeps}\rangle + \frac{1}{2}|\text{C: SB awaken}\rangle. \quad (6)$$

Thus, the measures Beauty should assign to A-branch and B-branch at t are 1/2 and 1/4 respectively. Similarly, she assigns 1/4 to C-branch at t'.

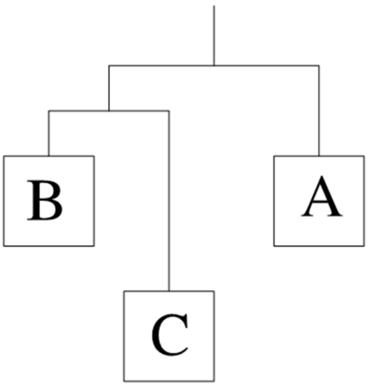

Figure 3. The alternative version of SB experiment with two coins.

What we strongly disagree about is putting the double-awakening in the original problem and two awakening following second coin toss in the two-coin version on equal footing. In other words, although the "illusion of probability" in a single quantum coin toss and Beauty's subjective degrees of belief in the simplified SB experiment are the same, it is a mistake to conclude that the double-awakening in the original problem can be replaced by a second coin toss.

In the simplified version, Lewis regards the discrepancy between Beauty's pre-branching certainty in "I will wake up on Monday/Tuesday" and between her post-branching uncertainty with respect to which day she is in. To reconcile between the two he appeals to the notion of genuine pre-branching uncertainty in the MWI and uses the structural parallel between the two scenarios to argue in favour of Beauty's pre-branching uncertainty in "I will wake up on Monday/Tuesday" (of 1/2) to bring it in line with her post-branching uncertainty in "Today is Monday/Tuesday". However, we believe it is perfectly legitimate for Beauty to be certain that she will be awaken on both Monday and Tuesday, while being uncertain about which day it is when she is awaken. This is fully consistent with the deterministic view advocated by us: there is no genuine uncertainty in the MWI. Lewis points out that even the pre-measurement certainty approach (i.e., the one advocated by us) incorporates a behaviour principle consistent with pre-measurement uncertainty. We agree. However, we do not believe that this supports his argument. The behaviour principle, associated with the measure, is based on the "illusion of probability", which means assigning uncertainty to the latter question – "Which branch am I in now?" – while prescribing credence of 1 for "Heads/Tails-branch will be realised".

Papineau and Durà-Vilà were the first to criticise Lewis's proposal to treat his version as equivalent to the original setup (Papineau and Durà-Vilà, 2009a). However, they also claimed that the version of the MWI with no genuine uncertainty strengthens Lewis's claim for 1/2, thereby questioning the legitimacy of our approach to probability in the MWI. There is a core disagreement between us and Papineau and Durà-Vilà. They argue against Lewis's conclusion for 1/2 in the full SBP by questioning the analogy between the quantum coin toss and the simplified SB case. As our approach strengthens this analogy, they conclude that it also strengthens Lewis's claim for 1/2. We have shown, however, that the analogy between the simplified SB and a quantum coin toss *does not* imply the equivalence between the two versions of the experiment. This refutes the claim by Papineau and Durà-Vilà that the approach to probability in the MWI advocated by us strengthens the claim for one half.

5. Summary and Outlook

We have analysed the SBP in the framework of the MWI. The main rationale for analysing problems like SBP within the MWI framework lies in the difficulties with conceptual interpretation of classical probability. As we have shown, within the MWI framework one can practice rational decision theory without appealing to the notion of classical probability. Instead, one can take advantage of the measure theory, which has a clear conceptual interpretation.

We have advocated the version of the MWI with no genuine uncertainty (Vaidman, 1998), were the observers' ignorance of future (or past) results of quantum experiments is recovered through the concept of the “illusion of probability”. We have reviewed the main concepts of this version of the MWI, such as the “measure of existence” (“caring measure”) and the “behaviour principle” and have compared this approach with *Subjective Uncertainty*, advocated by Saunders and Wallace. We have adopted the view that the solution to the SBP should not depend on the nature of the coin (quantum versus classical). We have emphasised that the question of assigning genuine pre-branching uncertainty in the MWI should not affect the solution to the Quantum SBP, for pre-branching uncertainty restores the standard (classical) decision theory considerations. Nevertheless, we believe that our approach to the solution is simpler and clearer than the Subjective Uncertainty approach.

We have also shown how the version of the MWI with no genuine uncertainty leads to the solution of the SBP consistent with a *Thirder* solution in the classical case. We have compared our approach with alternative views, especially with that of Peter Lewis, who argues for the answer 1/2, to which we do not agree.

Although we have focused our discussion on Peter Lewis's version of the SBP, there are several other recently published papers which analyse the SPB in the MWI. Daniel Peterson (2011), for example, claims that while the classical SBP yields the answer 1/3, in the Quantum version the answer should be 1/2. His version of the Quantum SBP involves a chain of events which is quite similar to Peter Lewis' two coins experiment. Darren Bradley (2011), using a variant of the SBP he called “The Technicolour Beauty”, also argues for the answer 1/2. Alastair Wilson (2013), on the other hand, criticises Bradley and argues for the answer 1/3. Our approach allows for simple calculation in all these cases, although this is beyond the scope of this paper.

Acknowledgements

B.G. thanks Sidney Sussex College, Cambridge for financial support. L.V. acknowledges the support of the Israel Science Foundation, Grant No. 1125/10.

References

Albert D, Loewer B. Interpreting the many worlds interpretation. Synthese 1988; 77: 195-213.

Bradley DJ. Confirmation in a branching world: The Everett interpretation and Sleeping Beauty. The British Journal for the Philosophy of Science 2011; 62: 323-342.

Elga A. Self-locating belief and the Sleeping Beauty problem. Analysis 2000; 60: 143-147.

Everett H. “Relative state” formulation of quantum mechanics. Rev Mod Phys 1957; 29:454-462.

Greaves H. Understanding Deutsch's probability in a deterministic multiverse. Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 2004; 35: 423-456.

Groisman B. The end of Sleeping Beauty's nightmare. The British Journal for the Philosophy of Science 2008; 59: 409-416.

Kent A. One world versus many: the inadequacy of Everettian accounts of evolution, probability, and scientific confirmation. In Saunders S, Barrett J, Kent A, editors, Many worlds?: Everett, quantum theory, and reality, chapter 10. Oxford University Press, 2010; pp. 307-354.

Lewis D. Sleeping Beauty: reply to Elga. Analysis 2001; 61: 171-76.

Lewis PJ. What is it like to be Schrödinger's cat? Analysis 2000; 60: 22-29.

Lewis PJ. Quantum Sleeping Beauty. Analysis 2007; 67: 59-65.

Lewis PJ. Reply to Papineau and Durà-Vilà. Analysis 2009; 69: 86-89.

Mensky MB. Reality in quantum mechanics, extended Everett concept, and consciousness. Optics and Spectroscopy 2007; 103: 461-467.

Papineau D, Durà-Vilà V. A thirder and an Everettian: a reply to Lewis's “quantum Sleeping Beauty”. Analysis 2009a; 69: 78-86.

Papineau D, Durà-Vilà V. Reply to Lewis: metaphysics versus epistemology. Analysis 2009b; 69: 89-91.

Peterson D. Qeauty and the books: a response to Lewis’ quantum Sleeping Beauty problem. Synthese 2011; 181: 367-374.

Saunders S. Time, quantum mechanics, and probability. Synthese 1998; 114: 373-404.

Saunders S. Chance in the Everett interpretation. In Saunders S, Barrett J, Kent A, editors, Many worlds?: Everett, quantum theory, and reality, chapter 6. Oxford University Press, 2010; pp. 181-205.

Saunders S, Wallace D. Branching and uncertainty. The British Journal for the Philosophy of Science 2008; 59: 293-305.

Tappenden P. Saunders and Wallace on Everett and Lewis. The British Journal for the Philosophy of Science 2008; 59: 307--314.

Tappenden P. Evidence and uncertainty in Everett’s multiverse. The British Journal for the Philosophy of Science 2011; 62: 99--123.

Vaidman L. On schizophrenic experiences of the neutron or why we should believe in the many-worlds interpretation of quantum theory. International Studies in the Philosophy of Science 1998; 12: 245-261.

Vaidman L. Probability and the many worlds interpretation of quantum theory. In Khrennikov A, editor, Quantum theory: Reconsideration of foundations. Section 2. Vaxjo University Press, 2001; pp. 408-412.

Vaidman L. Many-worlds interpretation of quantum mechanics. In Zalta EN, editor, The Stanford encyclopedia of philosophy. 2002; http://plato.stanford.edu/entries/qm-manyworlds.

Vaidman L. Probability in the many-worlds interpretation of quantum mechanics. In Ben-Menahem Y, Hemmo M, editors, Probability in physics. The Frontiers Collection, Springer Berlin Heidelberg, 2012; pp 299-311.

Vaidman L, Saunders S. On Sleeping Beauty controversy, 2001. Preprint, http://philsci-archive.pitt.edu/324/.

Wallace D. Epistemology quantized: circumstances in which we should come to believe in the Everett interpretation. The British Journal for the Philosophy of Science 2006; 57: 655-689.

Wilson A. Objective probability in Everettian quantum mechanics, 2012. Forthcoming in British Journal for the Philosophy of Science, doi:10.1093/bjps/axs022.

Wilson A. Everettian confirmation and Sleeping Beauty, 2013. Forthcoming in British Journal for the Philosophy of Science, http://alastairwilson.org/files/ecsbweb.pdf.