Failure of psychophysical supervenience in Everett's theory

Shan Gao[1,2]
[1] Research Center for Philosophy of Science and Technology,
Shanxi University, Taiyuan 030006, P. R. China
[2] Department of Philosophy, University of Chinese Academy of Sciences
Beijing 100049, P. R. China
E-mail: gaoshan2017@sxu.edu.cn.

August 7, 2017

Abstract

Psychophysical supervenience requires that the mental properties of a system cannot change without the change of its physical properties. For a system with many minds, the principle requires that the mental properties of each mind of the system cannot change without the change of the physical properties of the system. In this paper, I argue that Everett's theory seems to violate this principle of psychophysical supervenience. The violation results from the three key assumptions of the theory: (1) the completeness of the physical description by the wave function, (2) the linearity of the dynamics for the wave function, and (3) multiplicity. For a post-measurement state with two decoherent result branches, multiplicity means that each result branch corresponds to a mindful observer, whose mental properties supervene on the branch, and in particular, whose mental content contains a definite record corresponding to the result branch. Under certain unitary evolution which swaps the two result branches, the post-measurement state does not change, and the completeness of the physical description by the wave function then means that the physical state of the composite system does not change. While the linearity of the dynamics for the wave function requires that each result branch changes, and correspondingly the mental properties of the observer which supervene on the branch also change. Thus the principle of psychophysical supervenience as defined above is violated by Everett's theory.

Psychophysical supervenience is an important principle in the philosophy of mind. The standard definition.of supervenience is that a set of properties A supervenes on another set B in case no two things can differ with respect

to A-properties without also differing with respect to their B-properties (see McLaughlin and Bennett, 2014). By this definition, psychophysical supervenience requires that the mental properties of a system cannot change without the change of its physical properties.

Let me first give two examples which will be relevant to our later analayis of the status of psychophysical supervenience in Everett's theory. In the first example, a physical system has only one mind. The mental properties of the system can be formally represented by $M_1(C1)$, where the subscript index "1" denotes the identity of the mind, and $C1$ denotes the mental content of the mind. This simplified representation is enough for our later analysis. By this representation, the mental properties of the system can have three kinds of changes. The first kind of change is the change of mental content. It can be represented by

$$M_1(C1) \to M_1(C2). \tag{1}$$

The second kind of change is the change of identity. It can be represented by

$$M_1(C1) \to M_2(C1). \tag{2}$$

The third kind of change is the change of both identity and mental content. It can be represented by

$$M_1(C1) \to M_2(C2). \tag{3}$$

Obviously the third kind of change is greater than the first two kinds of changes. When the physical state of the system is unchanged, the principle of psychophysical supervenience requires that these changes of the mental properties of the system cannot happen, and the mental properties of the system should not change, namely

$$M_1(C1) \to M_1(C1). \tag{4}$$

In the second example, a physical system has two minds. The mental properties of the system can be formally represented by $M_1(C1)$ and $M_2(C2)$, where the subscript indexes "1" and "2" denote the identities of the two minds, respectively, and $C1$ and $C2$ denote the mental contents of the two minds, respectively. The principle of psychophysical supervenience requires that the mental properties of each mind of the system cannot change without the change of the physical state of the system. That is, when the physical state of the system is unchanged, only the following no-change of the mental properties of the system is permitted by the principle of psychophysical supervenience:

$$M_1(C1) \to M_1(C1); M_2(C2) \to M_2(C2). \tag{5}$$

For the purpose of facilating later analysis, let us consider two particular changes of the mental properties of the system. The first change is

$$M_1(C1) \to M_1(C2); M_2(C2) \to M_2(C1), \tag{6}$$

where each mind keeps her identity but swaps her mental content with the other. Obviously, after this change of each mind, the total mental properties of the system also change. The second change is

$$M_1(C1) \to M_2(C2); M_2(C2) \to M_1(C1), \tag{7}$$

where each mind not only swaps her mental content with the other but also swaps her identity with the other. Note that although this change is greater than the first change, the total mental properties of the system do not change after the change of each mind. However, since psychophysical supervenience requires that the mental properties of each mind of a system cannot change without the change of the physical state of the system, this change will also violate the principle of psychophysical supervenience if the physical state of the system is unchanged.[1]

After being familiar with the principle of psychophysical supervenience, let us turn to Everett's theory or the Everett interpretation of quantum mechanics. Everett's theory assumes that the wave function of a physical system is a complete description of the system, and the wave function always evolves in accord with the linear Schrödinger equation. In order to solve the measurement problem, the theory further assumes that after a measurement with many possible results there appear many equally real worlds, in each of which there is an observer who is aware of a definite result (Everett, 1957; DeWitt and Graham, 1973; Wallace, 2012). In the following, I will argue that Everett's theory seems to violate the principle of psychophysical supervenience as defined above.

Consider a simple spin measurement. First, suppose an observer O measures the x-spin of a spin one-half system S being x-spin up, $|up\rangle_S$.[2] By the Schrödinger equation, the physical state of the composite system after the measurement will evolve into the product state of O recording x-spin up and S being x-spin up:

$$|up\rangle_S |ready\rangle_O \to |up\rangle_S |up\rangle_O . \tag{8}$$

According to Everett's theory, there is still one observer, namely the original observer, after the measurement, and she is consciously aware of a definite record, x-spin up.

[1] One may object this conclusion. See later analysis.

[2] I will use the elegant Dirac notation throughout this paper.

Similarly, when the observer O measures the x-spin of a spin one-half system S being x-spin down, $|down\rangle_S$, the physical state of the composite system after the measurement will evolve into the product state of O recording x-spin down and S being x-spin down:

$$|up\rangle_S |ready\rangle_O \rightarrow |down\rangle_S |down\rangle_O . \tag{9}$$

Again, according to Everett's theory, there is still one observer, namely the original observer, after the measurement, and she is consciously aware of a definite record, x-spin down. It can be seen that these two mental evolution belongs to the first kind of mental change. The identity of the observer does not change, while her mental content changes.

Now consider a unitary time evolution operator, which changes $|up\rangle_S |up\rangle_O$ to $|down\rangle_S |down\rangle_O$ and $|down\rangle_S |down\rangle_O$ to $|up\rangle_S |up\rangle_O$, namely swaps the above two product states. It is similar to the NOT gate for a single q-bit, and is permitted by the Schrödinger equation in principle. Then after the evolution, the composite system being initially in the product state of O recording x-spin up and S being x-spin up will be in the product state of O recording x-spin down and S being x-spin down, namely

$$|up\rangle_S |up\rangle_O \rightarrow |down\rangle_S |down\rangle_O . \tag{10}$$

It seems that Everett's theory does not state whether the identity of the observer changes after the evolution. However, it is reasonable to assume (and I think many Everettians may also agree) that the identity of the observer does not change after the evolution, since the evolution only changes one conscious perception (and the relevant memory) of the observer, and it does not change all other mental properties of the observer, including all her previous memories. In this case, there is still one observer, namely the original observer, after the unitary time evolution, and her mental state changes from being aware of x-spin up to being aware of x-spin down.

Certainly, it may be also possible that the identity of the observer changes after the evolution according to a certain theory of identity (see Olson, 2017 for a review of theories of personal identity). In this case, the evolution should be more properly written as

$$|up\rangle_S |up\rangle_{O_u} \rightarrow |down\rangle_S |down\rangle_{O_d} . \tag{11}$$

where the identity of the observer changes from O_u to O_d, and her mental state also changes from being aware of x-spin up to being aware of x-spin down. As noted before, this mental change, which belongs to the third kind of mental change, is greater than the above mental change of (10), which belongs to the first kind of mental change.

Similarly, after the unitary time evolution, the composite system being initially in the product state of O recording x-spin down and S being x-spin

down will be in the product state of O recording x-spin up and S being x-spin up, namely

$$|down\rangle_S |down\rangle_O \rightarrow |up\rangle_S |up\rangle_O . \tag{12}$$

Correspondingly, there are also two possible mental evolution. In the first case, there is still one observer, namely the original observer, after the unitary time evolution, and her mental state changes from being aware of x-spin down to being aware of x-spin up. In the second case, the identity of the observer also changes from O_d to O_u besides this change of mental state, and the evolution may be more properly written as

$$|down\rangle_S |down\rangle_{O_d} \rightarrow |up\rangle_S |up\rangle_{O_u} . \tag{13}$$

These results are plain and familiar. Obviously the above evolution satisfies the principle of psychophysical supervenience. The mental properties of the composite system or the mental properties of the corresponding observer change with the change of the physical state of the composite system.

Let us consider a more interesting case. Suppose an observer O measures the x-spin of a spin one-half system S that is in a superposition of two different x-spins, $\frac{1}{\sqrt{2}}(|up\rangle_S + |down\rangle_S)$. By the linear Schrödinger equation, the physical state of the composite system after the measurement will evolve into the superposition of O recording x-spin up and S being x-spin up and O recording x-spin down and S being x-spin down:

$$\frac{1}{\sqrt{2}}(|up\rangle_S |up\rangle_O + |down\rangle_S |down\rangle_O). \tag{14}$$

According to Everett's theory, this post-measurement state corresponds to two observers, each of who is consciously aware of a definite record, either x-spin up or x-spin down.[3] Then, when considering the identities of the two observers, this post-measurement state may be written as

$$\frac{1}{\sqrt{2}}(|up\rangle_S |up\rangle_{O_u} + |down\rangle_S |down\rangle_{O_d}). \tag{15}$$

There are in general three ways of understanding the notion of multiplicity in Everett's theory: (1) measurements lead to multiple worlds at the fundamental level (DeWitt and Graham, 1973), (2) measurements lead to multiple worlds only at the non-fundamental "emergent" level (Wallace, 2012), and (3) measurements only lead to multiple minds (Zeh, 1981).[4] In

[3]Note that in Wallace's (2012) formulation of Everett's theory the number of the emergent observers after the measurement is not definite due to the imperfectness of decoherence. My following analysis also applies to this formulation.

[4]Note that Albert and Loewer's (1988) many-minds theory does not assume the usual notion of multiplicity as listed above. It assumes the existence of infinitely many minds even for a post-measurement product state, and it already entails dualism and violates the principle of psychophysical supervenience. I will not discuss this theory in this paper.

either case, for the above post-measurement state (15), the mental state of each observer is not determined uniquely by her whole wave function, but determined only by a branch of the wave function.[5]

Now consider again the above unitary time evolution operator, which changes the first branch of the superposition to its second branch and the second branch to the first branch:

$$\frac{1}{\sqrt{2}}(|up\rangle_S |up\rangle_O + |down\rangle_S |down\rangle_O) \rightarrow \frac{1}{\sqrt{2}}(|down\rangle_S |down\rangle_O + |up\rangle_S |up\rangle_O). \tag{16}$$

It can be seen that after the evolution the whole superposition does not change. According to Everett's theory, the wave function of a physical system is a complete description of the system. Therefore, after the above unitary time evolution the physical state of the composite system does not change.

On the other hand, as noted above, the above physical evolution has two possible corresponding mental evolution. In the first case, the identity of each observer does not change. Then the evolution will be

$$\frac{1}{\sqrt{2}}(|up\rangle_S |up\rangle_{O_u} + |down\rangle_S |down\rangle_{O_d}) \rightarrow \frac{1}{\sqrt{2}}(|down\rangle_S |down\rangle_{O_u} + |up\rangle_S |up\rangle_{O_d}). \tag{17}$$

In this case, like the corresponding product state case, after the evolution the mental state of each observer, which is determined by the corresponding branch of the superposition, will change; the mental state determined by the first branch will change from being aware of x-spin up to being aware of x-spin down, and the mental state determined by the second branch will change from being aware of x-spin down to being aware of x-spin up.[6] This is similar to the first change in the second example discussed above, namely (6). In the second case, the identity of each observer also changes besides the change of her mental state. Then the evolution will be

$$\frac{1}{\sqrt{2}}(|up\rangle_S |up\rangle_{O_u} + |down\rangle_S |down\rangle_{O_d}) \rightarrow \frac{1}{\sqrt{2}}(|down\rangle_S |down\rangle_{O_d} + |up\rangle_S |up\rangle_{O_u}). \tag{18}$$

This is similar to the second change in the second example discussed above, namely (7).

Therefore, Everett's theory predicts that after the above unitary time evolution the physical state of the composite system does not change, while

[5] It is worth noting that if the mental state of each observer is not determined by the corresponding branch of the post-measurement superposition, then the predictions of the theory will be not consistent with the predictions of quantum mechanics and experience for some unitary time evolution of the superposition.

[6] This is required by the linearity of dynamics. See below for further discussion.

the mental properties of the two involved observers both change after the evolution. According to the previous analysis of psychophysical supervenience, this means that Everett's theory violates the principle of psychophysical supervenience in the above example.

There are two possible ways to avoid the violation of psychophysical supervenience. The first way is to deny that after the evolution the physical state of the composite system has not changed. This requires that the wave function of a system is not a complete description of the physical state of the system, and additional variables are needed to introduce to describe the complete physical state. However, this requirement is not consistent with Everett's theory. Moreover, it is worth noting that in order to save psychophysical supervenience, it is also required that the additional variables should be changed by the unitary time evolution of the wave function, and the mental state of an observer should also supervene on the additional variables; otherwise the introduction of these variables cannot help save psychophysical supervenience in the above example.

The second way to avoid the violation of psychophysical supervenience in the above example is to deny that after the evolution the total mental states of the composite system have changed. This possibility deserves a more careful analysis. It seems uncontroversial that if the identity of each observer does not change during the above unitary time evolution (see (17)), then Everett's theory will violate the principle of psychophysical supervenience. If each observer has a trans-temporal identity and her mental state changes after the evolution, then the total mental properties of the composite system, which are composed of the mental properties of these observers, also change after the evolution. In this case, the total mental properties of the system are obviously different before and after the evolution (17).

However, it may be debatable whether Everett's theory really violates the principle of psychophysical supervenience when each observer also swaps her identity with the other after the above unitary time evolution (see (18)). The total mental properties of the composite system are the same before and after the evolution after all. If we define psychophysical supervenience for a system with many minds as the requirement that the mental properties of *each* mind cannot change without the change of the physical state of the system, then even if each observer swaps her identity with the other after the above evolution (18), Everett's theory also violates this requirement of psychophysical supervenience.

If one insists that Everett's theory does not violate the principle of psychophysical supervenience for the evolution (18), then one will meet at least two difficulties. The first one is that one must find a reasonable theory of identity to explain why the identity of an observer changes when all but one conscious perception (and the relevant memory) of the observer keeps unchanged. The second difficulty is that one must explain why and how the principle of psychophysical supervenience permits that when a system has

many observers the mental properties of each observer may change without the change of the physical state of the system. It is arguable that the second difficulty is much harder to solve than the first difficulty.

Here it may be worth noting that if one deny that each observer has her identity, then the violation of psychophysical supervenience can be avoided in the above example. The reason is that after the above evolution there remain a mental state corresponding to seeing a spin up result and a mental state corresponding to seeing a spin down result, and thus the total mental states of the composite system have not changed. However, the absence of identities of observers seems obviously inconsistent with our experience. Moreover, if each observer has no identity in the above example, then this seems equivalent to say that there is only one observer with two mental contents, which are seeing a spin up result and seeing a spin down result (see Gao, 2016, 2017 for further discussion). This is not consistent with Everett's theory.

In order to avoid the violation of psychophysical supervenience, one may even assume that after the above unitary time evolution the identity and mental state of each observer are not changed. However, the mental dynamics must also keep the identity and mental state of an observer being in one branch of the above superposition (14) unchanged so that the mental dynamics is still linear.[7] This seems inconsistent with the predictions of quantum mechanics and experience. On the other hand, if the mental dynamics still changes the mental state of the observer being in one branch of the above superposition (14) as usual, then the mental dynamics must be nonlinear. Although a nonlinear dynamics for the physical state or the wave function is obviously inconsistent with Everett's theory, a nonlinear dynamics for the mental state seems to be not prohibited by the theory; the many-minds theory is an example (Albert and Loewer, 1988; Barrett, 1999). However, the existence of a nonlinear dynamics for the mental state in Everett's theory already entails dualism. It seems that this is no better than the violation of psychophysical supervenience. Moreover, such a nonlinear dynamics seems very ad hoc, and it is also difficult to determine what the dynamics is for an arbitrary superposition such as $\alpha \left|up\right\rangle_S \left|up\right\rangle_O + \beta \left|down\right\rangle_S \left|down\right\rangle_O$, where α and β are not zero and satisfy the normalization condition $|\alpha|^2 + |\beta|^2 = 1$.

Finally, one may argue that the above superposition (14) is a very special state, and thus the violation of psychophysical supervenience, even if it

[7] A linear dynamics requires that the evolution of one branch of a superposition is independent of the evolution of other branches, as well as whether or not these branches exist. Thus, by the same unitary evolution operator, the evolution of one branch of the post-measurement superposition (14), such as the branch $\left|up\right\rangle_S \left|up\right\rangle_O$ in the sperposition, will be the same as the evolution of the post-measurement state containing only this branch, such as the product state $\left|up\right\rangle_S \left|up\right\rangle_O$. This is true for the evolution of both the physical state and the mental state. Otherwise the linearity of dynamics will be violated, and the resulting dynamics will be nonlinear.

exists, is not serious for Everett's theory. For other states, the amplitudes of the two branches of the superposition are different, and thus after the above evolution the physical state of the composite system, like the mental states of the system, will also change. Then the psychophysical supervenience will not be violated for these states. However, for a general superposition we can use an additional unitary time evolution operator, which changes $\alpha \left|up\right\rangle_S \left|up\right\rangle_O + \beta \left|down\right\rangle_S \left|down\right\rangle_O$ to $\beta \left|up\right\rangle_S \left|up\right\rangle_O + \alpha \left|down\right\rangle_S \left|down\right\rangle_O$, besides the above unitary time evolution operator. Then by the combination of the two unitary time evolution operators which is still unitary, the superposition may also keep unchanged. But, by a similar analysis as above, the mental properties of the system change after the evolution.

In addition, even if using only the original unitary time evolution operator, it seems that there is still a potential problem with the realization of psychophysical supervenience for some other states. When the difference of the amplitudes of the two branches of a post-measurement superposition is very small, the change of the physical state of the composite system is also very small after the evolution. But the change of the mental state of each observer is still very large, e.g. from being aware of x-spin down to being aware of x-spin up. In this case, although the psychophysical supervenience is not violated in a strict sense, it seems very difficult or even impossible to explain how the mental state supervenes on the physical state.

To sum up, I have argued that Everett's theory seems to violate the principle of psychophysical supervenience. The violation results from the three key assumptions of the theory: (1) the completeness of the physical description by the wave function, (2) the linearity of the dynamics for the wave function, and (3) multiplicity. It seems that one must go beyond Everett's theory in order to avoid the violation of psychophysical supervenience.

Acknowledgments

This work is partly supported by a research project grant from Chinese Academy of Sciences and the National Social Science Foundation of China (Grant No. 16BZX021).

References

[1] Albert, D. Z. and B. Loewer. (1988). Interpreting the Many Worlds Interpretation, Synthese, 77, 195-213.

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

[3] DeWitt, B. S. and N. Graham (eds.). (1973). The Many-Worlds Interpretation of Quantum Mechanics. Princeton: Princeton University

Press.

[4] Everett, H. (1957). 'Relative state' formulation of quantum mechanics. Rev. Mod. Phys. 29, 454-462.

[5] Gao, S. (2016). What does it feel like to be in a quantum superposition? http://philsci-archive.pitt.edu/11811/.

[6] Gao, S. (2017). The Meaning of the Wave Function: In Search of the Ontology of Quantum Mechanics. Cambridge: Cambridge University Press.

[7] McLaughlin, B. and Bennett, K. (2014). Supervenience, The Stanford Encyclopedia of Philosophy (Spring 2014 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/spr2014/entries/supervenience/.

[8] Olson, E. T. (2017). Personal Identity, The Stanford Encyclopedia of Philosophy (Summer 2017 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/sum2017/entries/identity-personal/.

[9] Wallace, D. (2012). The Emergent Multiverse: Quantum Theory according to the Everett Interpretation. Oxford: Oxford University Press.

[10] Zeh, H. D. (1981). The Problem of Conscious Observation in Quantum Mechanical Description, Epistemological Letters of the Ferdinand-Gonseth Association in Biel (Switzerland), 63. Also Published in Foundations of Physics Letters. 13 (2000) 221-233.