The role of the observer in the Everett interpretation

H. D. Zeh – www.zeh-hd.de (Oct. 2012)

Abstract: The role attributed to the observer in various interpretations of quantum mechanics as well as in classical statistical mechanics is discussed, with particular attention being paid to the Everett interpretation.

Keywords: quantum measurement, decoherence, physics of observation, incomplete information, meaning of ensembles

1. The observer in traditional quantum mechanics

When Werner Heisenberg discovered his matrix mechanics, which denies the existence of definite values for classical physical variables such as position and momentum of an electron before they are measured, he invented a historically unprecedented role for "human observers". He assumed that reality is *created* in an irreversible process of observation performed by humans – for him confirmation of the superiority of an idealistic world view instead of materialism, realism and reductionism. In particular, the concept of time (including its arrow) would become a fundamental extra-physical prerequisite for the formulation of this process as well as of other physical laws. This point of view was soon supported by his friends Wolfgang Pauli and Carl Friedrich von Weizsäcker. It seemed to become even strengthened when Heisenberg's early attempts to understand his uncertainty principle simply as a consequence of unavoidable perturbations of the electron during measurements (for example by his "electron microscope") failed as a consistent explanation.

Niels Bohr later subscribed to a somewhat different position by assuming that the outcome of a measurement is *objectively* created as a classically described property in the measurement apparatus. However, he rejected all attempts to analyze this measurement as a dynamical physical process, since he also denied any observer-independent microphysical reality in order to avoid otherwise apparently arising consistency problems. So he insisted that we cannot speak about nature herself, but only about what *we* (humans) can know about her.

Erwin Schrödinger expected initially that the single particle wave function he had postulated describes a real electron, and thus explains the uncertainty between position and momentum by means of the Fourier theorem, but he could not understand the apparently required quantum jumps between the standing waves that represent energy eigenstates. For a long time he tried to get around the formally arising consequence of entangled many-particle wave functions or the entanglement of spatially separated systems – although precisely this entanglement offers an explanation of the apparent jumps (as we know today). When Max Born invented his interpretation of wave functions as probability amplitudes for classical particle properties, this entanglement was insufficiently explained as representing no more than probability correlations. So Heisenberg later spoke somewhat mystically about the wave function as representing "human knowledge as an intermediary kind of reality" – an idea that seems to have been revived in the recently quite popular information theoretical approach to quantum theory.

Meanwhile, John von Neumann (in his book about the Mathematical Foundations of Quantum Mechanics) studied the possibility of describing the measurement process as a quantum mechanical interaction between the object and the apparatus. In strong contrast to Bohr's view, he represented states of the macroscopic "pointer" themselves by wave packets rather than in classical terms. The unitary interaction, when applied to an initial microscopic superposition, would then lead to superpositions of different pointer positions entangled with different positions of the measured particle, for example. Therefore, von Neumann had to postulate a stochastic collapse (or "reduction") of the wave function as a new kind of dynamics supplementing the Schrödinger equation. He called it a "first intervention" – probably since at that time mainly energy eigenstates were studied, while quantum dynamics appeared to consist of quantum jumps between them. The time-dependent Schrödinger equation (his "second intervention") was mostly used to calculate probabilities for such jumps. Actually, the collapse came close to Born's first version of his probability interpretation, which postulated transitions between initial and final wave functions. It was only later re-interpreted by Pauli as describing probabilities for the occurrence of classical (such as particle) properties.

However, von Neumann did not stop his quantum dynamical considerations with the apparatus. He also included the observer himself as a quantum system. In this way, the collapse was important for him in order to re-establish a "psycho-physical parallelism", which would not have been possible if the observer could be in a superposition of physical states representing different states of awareness – similar to Schrödinger's cat being simultaneously dead and alive. In order to achieve this goal, the collapse or any other probability interpretation could be equally applied at each step of the "indivisible chain of interactions between the observer and the observed" (Weizsäcker's words). This freedom was also referred to as the variability of the "Heisenberg cut". For Heisenberg, it was a fundamental element in his understanding of quantum theory,[1] while Bohr preferred to apply it at some not very precisely specified border line separating the microscopic and the macroscopic world within the measurement apparatus.

The conscious observer was further discussed as the key element of the quantum measurement process by London and Bauer.[2] Eugene Wigner later even suggested explicitly the possibility of an active influence of consciousness on the physical world,[3] but dropped this proposal when he learned about the concept of decoherence. Experiments to confirm such effects in the form of deviations from Born's rule, caused by the observer's mind, have failed, however. Quantum indeterminism thus seems to have nothing to do with an apparent "free will". While I therefore prefer to understand von Neumann's "psycho" part in his parallelism in the sense of a mere epiphenomenon accompanying a passive physical part, Max Jammer compared it with Anaxagoras' dualistic doctrine of Matter and Mind when he quoted (adding his suggested modern interpretation in parentheses):[4] "The things that are in a single world are not parted from one another, not cut away with an axe, neither the warm from the cold nor the cold from the warm" (superpositions!?), but "when Mind began to set things in motion, separation took place from each thing that was being moved, and all that Mind moved was separated" (reduction!?).

Other interpretations of quantum mechanics, such as Bohm's, are often claimed *not* to require an observer. Although the observer does indeed not assume a specific role in the dynamics of this theory, John Bell pointed out that Bohm's theory is tacitly based on the assumption of an observer being physically described solely by the "classical" and thus local variables that are here simply assumed to exist in addition to the non-local wave function.[5] Collapse theories, on the other hand, would not only have to explain the occurrence of definite narrow wave packets for all macroscopic variables, but also definite states of the conscious observer system if one wants to eliminate (rather than merely decohere) superpositions of different states of awareness.

2. The observer in classical statistical physics

The observer has always played an essential role in the empirical sciences, simply because they are based precisely on observations performed by humans by physical means. This remark may appear trivial, but its consequences are non-trivial for all physical concepts that depend on "incomplete information", such as in statistical mechanics.[6] Why do we regard the position and shape of a solid body as "physically given" even when we do not know them, while we describe its molecules objectively by a distribution of states characterized by a certain temperature, for example? Since in a Laplacean world all variables are equally real, any such distinction must be based on the difference between what *we* can easily observe and what would require a certain instrumental effort to find out. Such differences, often based on arguments of dynamical stability versus rapid change, are certainly relevant, but the boundary separating their realms may even vary, for example when we decide to take into account local fluctuations of certain equilibrium values, or more drastically in phase transitions.

This dependence of physical concepts on incomplete knowledge is particularly obvious in Willard Gibbs' approach to statistical thermodynamics, which is based on Γ-space distributions and appears related to general concepts first proposed by Thomas Bayes. Boltzmann's μ-space distributions, on the other hand, seem to represent reality rather than information. However, they would correspond to a complete description (an individual *point* in Γ-space) if we did not neglect some information, for example by applying a coarse graining or smoothing of the real (discrete) distribution in μ-space, and if particles were distinguished from one another according to their trajectories. So one may already raise John Bell's fundamental question "Information by whom?", while his additional "about what?" would here be answered as "about points in Γ-space", although this answer fails to explain the empirically required absence of the factorials $N!$ that would result from particle permutations.

While the physically important entropy concept can then be statistically defined by the size of the ensemble of microscopic states representing a macroscopic situation (or, equivalently, by their mean probability in such an ensemble), the precise definition of this ensemble is in general not very relevant. Since entropy is a logarithmic measure, an uncertainty by a factor of X in the size of the ensemble in Γ-space would give rise to a relative correction merely of the order $\ln X/10^{20}$. Therefore, one may readily interchange slightly different "representative ensembles" without doing much harm. Precision of concepts does become essential, however, for questions of principle, such as in measurements or in considerations regarding

Maxwell's demon or Szilard's engine. Even the difference between a canonical and a microcanonical ensemble is physically meaningful, as only the former contains lacking information about the precise energy, such as caused by fluctuations in an open system.

In particular, microscopic determinism requires the dynamical conservation of the size of an arbitrary ensemble (its formal ensemble entropy), whatever its origin. However, if physical entropy is defined as a function of given macroscopic properties, it is at most indirectly related to an ensemble that represents information held by an observer. For example, physical entropy is an extensive (additive) quantity, while ensemble entropy is *not*, since it would strongly depend on dynamically arising probability correlations between subsystems. This is the reason while Boltzmann's μ-space entropy may change in time even for deterministic collisions between the particles. More importantly, in irreversible phase transitions, new macroscopic variables (such as droplet positions in a condensation process) or new order parameters are often created. In this case, lacking information about thermal degrees of freedom (physical entropy) is transformed into lacking information about macroscopic variables (that is, information entropy in the original sense of Claude Shannon). Hence, physical entropy *can* be lowered if we regard macroscopic properties as "given" as soon as they arise. Wile this is essentially a matter of definition of physical entropy, the true surprise comes when one adds the (human?) observer to the chain of interacting systems in analogy to von Neumann's description of observations following quantum measurements. Deterministically, different macroscopic properties are thereby correlated with different states of the observer. If the observer is assumed to know what he has observed, the ensemble describing his lack of knowledge would be reduced without violating microscopic determinism, in this way reducing the initial ensemble entropy! One may even conceive of a classical version of *Wigner's friend* for this purpose. (In all these examples, the creation of *uncontrollable* correlations by molecular collisions, which is the major source of irreversibility, has been disregarded for simplicity.)

The situation described above is related to Maxwell's demon, who was proposed to use his presumed initial knowledge about molecular motions in order to reduce thermal entropy. Leo Szilard demonstrated by means of a thought experiment that the demon's entropy must correspondingly increase if he is regarded as a physical object. (In Rolf Landauer's language: information is physical.) Charles Bennett therefore concluded[7] in accordance with traditional formulations of the second law that Maxwell's demon cannot work in a cyclic process as required for a perpetuum mobile of the second kind, because he would have to get rid of his entropy in order to close the cycle. However, lowering the ensemble entropy in an *indi-*

vidual process as described above would nonetheless be possible in principle by interaction with a participating observer. On the other hand, any observer must have got rid of quite a lot of entropy in order to come into being in a process of self-organization.

From a thermodynamical point of view, the (neg)entropy of information is usually neglibible. Therefore, observers are often regarded as extra-physical, or as possessing infinite information capacity and even defining their own arrow of time. This position becomes problematic, in particular, when applied to quantum states and assuming that the latter represent a novel concept of extra-physical "quantum information" that is not counted in any definition of entropy. Any measure of information must presume an ensemble of possibilities to which this information applies. In classical context, all measurement outcomes are in principle determined in advance by the global microscopic state, and may thus be "measured" (that is, selected rather than created). While microscopic variables can usually be regarded as randomized before being measured, macroscopic ones are redundantly "documented" in their environment (for example by the light that they have scattered and that might later be received by observers). For this reason they appear to be "objectively given"; one cannot conceive of *one* individual document only being changed in order to change the past. This retarded documentation of macroscopic "facts" requires a strong time asymmetry of the physical world that establishes its "causal appearance" and lets the macroscopic past appear fixed.

3. The observer in the Everett interpretation

Hugh Everett first recognized that we don't have to postulate a dynamical collapse of the wave function if we accept instead that the subjective observer may exist in various "versions" *i* that result from von Neumann's unitary description of a quantum observation

$$(1) \qquad \left(\sum\nolimits_i c_i \psi_i^S\right)\psi_0^A \psi_0^O \rightarrow \left(\sum\nolimits_i c_i \psi_i^S \psi_i^A\right)\psi_0^O \rightarrow \sum\nolimits_i c_i \psi_i^S \psi_i^A \psi_i^O =: \sum\nolimits_i c_i \psi_i^{rel} \psi_i^O \quad .$$

Its first step is sometimes called a "pre-measurement". Here the suffixes *S, A, O* indicate the system, apparatus and observer, respectively. For simplicity, any information medium, such as light, is regarded as part of the apparatus. The states ψ_i^{rel} on the right hand side, which are identical to the potential states resulting from a stochastic collapse, define the "relative state" of the outside world with respect to the physical state of the subjective observer ψ_i^O (therefore the title "Relative State Interpretation" of Everett's original publication). The observer, although remaining passive in (1), evidently assumes a crucial role in Everett's description.[8,9,10]

Emphasis on this aspect has also led to the name "many minds" or "multi-consciousness" interpretation. (See also the explicit examples of complete observations at the end of this section.)

Everett's conclusion was almost unanimously rejected by the physics community at its time for several reasons (if not just emotionally because of its unconventional nature). The major one was that the wave function had traditionally been regarded as meaningful only in the microscopic world, or at most until the Heisenberg cut is applied, whereupon it was assumed to "lose its meaning". Those who did consider the general validity of the wave function as a possibility raised another objection: the expansion $\psi^{total} = \sum_i \psi_i^{rel} \psi_i^O$ is defined with respect to any basis ψ_i^O chosen for the observer system (including states which represent superpositions of different states of awareness) unless their relative states, too, were required to be mutually orthogonal. In von Neumann's equation (1), the *i*-basis is usually defined by a phenomenological "observable" that is used to characterize a measurement, no matter whether an observer ever entered the scene in order to read off the result. This basis was later called the "pointer representation". If the observer system *O* itself were precisely defined (as one might hope for a minimum system representing consciousness), one could use the essentially unique Schmidt canonical representation, in which *both* subsystem bases are exactly orthogonal. However, this representation would fluctuate in time and strongly depend on the precise boundary between observer system *O* that seems to form the physical side of a psychophysical parallelism and the rest of the quantum world.[11] It is thus not appropriate for a description of objective measurements by an apparatus.

The problem of how to define an objective pointer basis that is sufficient for all practical purposes of observers was resolved by the theory of environmental decoherence.[12] Accordingly, the relative states in (1) have to contain, in an essential way, an uncontrollable and normally inaccessible environment of the macroscopic apparatus,

$$(2) \qquad \psi_i^{rel} = \psi_i^S \psi_i^A \psi_i^{env} \quad .$$

Because of the unavoidably arising *i*-dependence of the environmental states ψ_i^{env}, a superposition of macroscopically different states ψ_i^A formed in a measurement is immediately and irreversibly dislocalized. In this way, the "normal environment" of a macroscopic system usually induces a preferred basis for the pointer variable or other quasi-classical property that is objectively characterized by its rebustness against further decoherence. This decoherence phenomenon was the first clear indication that entangled wave functions are in fact valid and meaningful beyond closed microscopic systems, although this entanglement must be far more complex than envisioned by von Neumann and Everett with their simplified model. The *i*-dependent effect in the environment need *not* represent any useful information or documentation, since a thermal environment suffices to achieve decoherence. Therefore, only a minority of quasi-classical variables may be assumed to be "always given".

Because of the locality of all interactions, the preferred basis is usually the position basis of a pointer or other macroscopic variable. Although a reversal of this dislocalization would in principle be compatible with the Schrödinger equation, it is excluded by the arrow of time characterizing our world (namely the absence of any advanced or "anti-causal" correlation or entanglement that would have local effects in the future). The nonlocal superpositions that are "caused" according to (1) when taking into account the environment can then not be relocalized. Since, on the other hand, they cannot disappear from the universe unless the Schrödinger dynamics would have to be changed, any description of reality in terms of the unitarily evolving wave function requires an Everett interpretation.

Realistically, the macroscopic "apparatus" in (1) would not only have to include any required information medium or registration device, but even the human sensory organs and most of the neuronal system. Both are macroscopic in the sense of being decohered by their environments,[13] and both are external to any reasonable subjective observer system (the final carrier of consciousness). The neuronal apparatus is indeed a particularly fine-grained quasi-classical system, whose variables can be assumed to be always "given". However, this further decoherence does not affect the measurement proper as an objective physical process. The precise "localization" of consciousness in the physical world remains an open problem - similarly as it did in classical descriptions, although one may expect that it has ultimately to be defined in quantum mechanical terms.

After decoherence by the environment, the macroscopic system may for all practical purposes be described by its reduced density matrix, such as $\rho_{red}^{A} = \sum_i \left|c_i\right|^2 \overset{*}{\psi_i^A}\ \psi_i^A$. Although this density matrix is identical to the one that would represent the ensemble of states postulated by a collapse, it does here *not* represent an ensemble. However, the global superposition (1) – including the environment – now consists of various corresponding autonomous world components which describe different macroscopic properties. Therefore, the different observer states ψ_i^O, whatever their precise definition, cannot dynamically feel the presence of the “other worlds” that are described by the states $\psi^{rel}_{i'\neq i}$ any more. This consequence is sufficient for the theory to consistently describe all our observations in an apparently classical universe. Note that the molecules forming a gas are also decohered into narrow wave packets by their mutual collisions, and thus approximately define a quasi-classical μ-space distribution, but this does not justify quasi-deterministic trajectories for them. Their collisions would appear fundamentally stochastic in such a quasi-classical description. For this reason, their positions, although decohered, cannot be assumed to be “given” in an Everett branch world.

After the dynamical autonomy of Everett branches has been clearly established by decoherence, the localization of the subjective observer system in states existing within these branches appears indeed so plausible or “normal” that, for example, the Oxford quantum philosophers regard the quantum measurement problem as solved by the combination decoherence plus Everett without explicitly mentioning the observer – again in analogy to the classical concept of an observer-independent reality.[14]

As an example, consider two spatially separated microscopic systems entangled with one another as in a Bell experiment. If one of them is locally measured, *both* get immediately entangled with the apparatus and its local environment – nothing else as yet. An observer at the location of the other microscopic system will participate in the entanglement only after having received a signal about the outcome. Only thereafter will he be in different states in the different “worlds” that were dynamically defined by the irreversible decoherence, but which together still form but *one* quantum world. If he decides to measure and observe also the other microscopic system (before or after receiving the first signal), his state has to split further in order to register and to be aware of both quasi-classical outcomes. When repeating the total measurement several times, he would in “most” of his versions in very good approximation confirm the frequencies predicted by Born’s rule and their correlations that violate Bell’s inequality – provided the branches containing his various versions are *assumed* to possess statistical weights according to their squared norms when defining what is meant by “most”. Any other conceivable statistical weights that are *not* conserved under the Schrödinger equation (such as the ill-defined *number* of branches) would lead to probabilities that might later change under further measurements that are asymmetrically performed in different branches. Everett considered this consequence as proof of Born’s probability measure in terms of the squared norm.[15]

All those “weird consequences” of quantum mechanics that have recently been “discovered” and subsequently much discussed in the media can be similarly described consistently in wave mechanical terms, since this is precisely the way how they were all predicted. Their apparent weirdness is merely a consequence of the traditional interpretation of the wave function as representing no more than probabilities for locally defined classical quantities. The reader may himself analyze the so-called quantum teleportation protocol as a second example in order to confirm that nothing is (apparently) teleported that had not been prepared in advance at its target position in one of the components of the required entangled wave function.[16] Teleportation and other “esoteric” phenomena would only be required if local properties were *created* in measurements (as assumed in the Copenhagen interpretation). It is evident that entanglement cannot represent “nothing but information”, even though one may *pretend* that an ensemble representing lacking information is created when decoherence first becomes irreversible in a chain of interactions that leads to observation. (The cat has to be assumed to have died – if it ever did in the corresponding “world” – long before the box was opened.) As this apparent collapse is not a *physical* process, it may even be defined to “occur” superluminally. This restriction of *apparent quantum reality* to one effective branch wave function at the time of first decoherence is certainly convenient and thus pragmatically justified, but physics students should be expected to understand its origin in a complete and consistent theory (as just described).

Any proposal for a *genuine* collapse would have to be specified in order to be meaningful, but thereby has to avoid consistency problems with the principles of relativity. On the other hand, any conceivable confirmation of such a violation of the Schrödinger equation would readily *falsify* Everett’s interpretation, while an unspecified collapse proposal can hardly ever be falsified. Most collapse models contain free parameters that would also render them non-falsifiable as long as these parameters do not have to violate certain bounds that are required for them to fulfill their purpose of predicting definite measurement outcomes. Therefore, the dispute about a collapse of the wave function versus Many Worlds is not a matter of belief or religion. There simply exist two classes of possibilities whose consequences should be further analyzed and tested, while the original Copenhagen interpretation with its fundamental classical concepts seems by now to be deprived of all motivation by the success of the decoherence program. It should also be obvious that the wave function can carry information only if it is a physical (real) object.

The consistent description of quantum phenomena according to Everett means that the observed quantum indeterminism does not correspond to a stochastic dynamical process in nature, since the global wave function is assumed to evolve deterministically. Rather, it reflects the multiple future of an observer in this deterministic quantum world – comparable to a process of cell division in a classical world that could thereby remain deterministic. However, this indeterminism of the observer is objectivized with respect to those versions of different observers (including those of “Wigner’s friend”) that are correlated by their entanglement. Their versions who “live” in the same world branch always agree about the outcome of measurements, while other versions do not have to disappear from reality; it is sufficient that they cannot communicate any more with one another. This entanglement between different observers is the same as that between an observer and his apparatus.

Acknowledgement

I wish to thank Erich Joos for his critical revision of this manuscript.

4. References

---

[1] G. Bacciagaluppi and E. Crull, Studies in Hist. Philos. Mod. Physics **40**, 374 (2009)

[2] F. London and E. Bauer, *La théorie de l'observation en méchanique quantique* (Hermann, Paris 1932)

[3] E. Wigner, in *The Scientist Speculates*, L.G. Good, edt. (Heinemann, London 1962), p. 284

[4] M. Jammer, *The Philosophy of Quantum Mechanics* (Wiley, New York 1974), p. 482

[5] J.S. Bell, in *Quantum Gravity 2,* C.J. Isham, R. Penrose, and D.W. Sciama, edts. (Oxford UP 1981)

[6] H. D. Zeh, *The Physical Basis of the Direction of Time* (Springer, Heidelberg 2007)

[7] C.H. Bennett, Scient. Am. **253**, 88 (1957)

[8] H.D. Zeh, Found. Phys. **13**, 221 (2000)

[9] M. Menskii, Physics - Uspekhi **48**, 389 (2005)

[10] M. Tegmark, Phys. Rev. **D 85**, 123517 (2012)

[11] O. Kübler and H.D. Zeh, Ann. Phys. (N.Y.) **76**, 405 (1973); H.D. Zeh, Found. Phys. **9**, 803 (1979)

[12] H.D. Zeh, Found. Phys. **1**, 69 (1970); E. Joos, H.D. Zeh, C. Kiefer, D. Giulini, J. Kupsch and I.-O. Stamatescu, *Decoherence and the Appearance of a Classical World in Quantum Theory,* (Springer, Heidelberg 2003), Chpts. 1-4

[13] M. Tegmark, Phys. Rev. **E61**, 4194 (2000)

[14] See the contributions by S. Saunders and D. Wallace in *Many Worlds?,* S. Saunders, J. Barrett, A. Kent, and D. Wallace, edts. (Oxford UP 2010)

[15] P. Byrne, *The Many Worlds of Hugh Everett III* (Oxford UP 2010)

[16] H.D. Zeh, Found. Phys. **40**, 1476 (2010)