Erkenntnis **48** (1998), 39–46

Ignorance of the Own Past

Thomas Breuer
Institut für Philosophie, Universität Salzburg
Franziskanergasse 1, A-5020 Salzburg

Abstract

It is shown that for an inside observer it is impossible to distinguish all states in which a system was at some past time. This holds for classical and quantum systems, but an assumption of determinism is essential in the proof.

1 Introduction

One cannot know about oneself that one knows something different; thus one cannot know everything about oneself. This is the way popular arguments go claiming that complete self-knowledge is impossible. But one can know about oneself that one *knew* something different; thus one can perhaps know everything about the own past, at least in principle. Contrary to what is suggested by this I will argue that complete self-knowledge about the past is impossible even in principle, at least for deterministic time evolutions.

Let us speak physics instead of information theory. There have been suggestions[1] that physical systems look different to inside observers than they do to outside observers. Accordingly it is claimed that the physics of an inside observer (endophysics) differs from the physics of an outside observer (exophysics). For inside observers problems of self-reference become relevant because measurements from inside are self-measurements.[2]

[1] See e.g. Finkelstein (1988), the papers in Kampis et al. (1993), Mittelstaedt (1993), Peres and Zurek (1982), Penrose (1989), Popper (1950), Rössler (1987), Svozil (1993).

[2] Whether the quantum meachanical measurement problem is a problem is self-reference—as indicated by Dalla Chiara (1977), Peres and Zurek (1982), Penrose (1989), Primas (1990)—is investigated in my (1996).

In my (1995) I demonstrate restrictions on measurability from inside which may justify such a distinction both for classical and quantum mechanical systems. These results on self-referential measurements were obtained in a static context: for an observer it is impossible to distinguish all *present* states of a system in which he is contained. Here I extend these results to a dynamical context; the question is whether an inside observer can distinguish all states a system was in at some *past* time. It turns out that this is impossible provided the time evolution is deterministic. I do not know whether a similar result holds for stochastic time evolutions.

2 Description of measurements

Assume an observer A is performing a measurement on a system O. For the time being A may or may not be contained in O. The measurement is an interaction between A and O which establishes a relation between the states of A and those of O. From information about the state of A at a certain time t_1 after the measurement we infer information about the state of O at an earlier time t_0.

This inference can be described by an inference map $I : P(S_A^1) \to P(S_O^0)$ from the power set $P(S_A^1)$ of the state space S_A^1 of the apparatus at t_1 to the power set $P(S_O^0)$ of the state space S_O^0 of O at t_0. (The upper index of S_O^0 or S_1^A refers to the time, the lower to the system. By introducing the upper index I do not want to imply that there are many interesting cases where $S_O^0 \neq S_O^1$ or $S_A^1 \neq S_A^0$. It just makes it easier to indicate to which times states and sets of states refer.) I is defined in the following way:

> I associates to every set $X \subset S_A^1$ of apparatus states at time t_1 the set $I(X)$ of those object states at time t_0 which are compatible with the information that at t_1 the apparatus is in one of the states in X.

I depends on the kind of experiment being performed. When the experimenter chooses the experimental set-up he also decides how he is going to interpret the pointer readings. This is described by I.

I is defined as a map of sets of states and not as a map of states for the following reasons: firstly, after the measurement the experimenter usually does not know the exact state of the apparatus but only the pointer reading; secondly experiments in general do not determine exactly one state of the observed system but rather the (perhaps approximate) value of a phys-

ical quantity. A measurements of some quantity can also be regarded as measurement of a set of states.

At this point let me mention a consequence of the definition of I which will be used later. For every set $X \subset S_A^1$ of apparatus states we have

$$I(X) = \bigcup_{s \in X} I(\{s\}) \tag{1}$$

This is easy to see. Since $I(\{s\})$ is the set of object states at t_0 compatible with the apparatus state at t_1 being s, $\bigcup_{s \in X} I(\{s\})$ is the set of object states at t_0 compatible with the apparatus state at t_1 being in some state in X. By definition this is $I(X)$. (To simplify notation from now on I will write $I(s)$ instead of $I(\{s\})$.)

Some more definitions. An experiment with inference map I is said to be able to *measure a state s of O at t_0 exactly* if there is a set X_s of apparatus states at t_1 referring uniquely to s: $I(X_s) = \{s\}$. An experiment with inference map I can *distinguish two object states* s_1, s_2 if there is one set X_1 of apparatus states at t_1 referring to s_1 but not to s_2 and another set X_2 referring to s_2 but not to s_1: $I(X_1) \ni s_1 \notin I(X_2)$ and $I(X_1) \not\ni s_2 \in I(X_2)$.

I have to assume that the time evolution of O from t_0 to t_1 is *deterministic*:

> The state of O at t_0 determines uniquely the state at t_1.

This implies that there is a map $T : S_O^0 \to S_O^1$ associating to every state $s \in S_O^0$ the state $T(s) \in S_O^1$ into which s evolves. T is surjective because every possible state of O at t_1 must have evolved from some possible state at t_0. T can be extended to the power set $P(S_O^0)$ by defining $T(X) := \bigcup_{s \in X} T(s)$. For the extended map I will use the same letter T.

Actually, in the standard literature[3] the term "determinism" is only applied if the state of the world at one time determines the state at *all* future and past times. Here determinism is assumed only for the time evolution between t_0 and t_1. This assumption does not imply that the time evolution has to be always deterministic. Therefore the assumption does not exclude the freedom of the experimenter to choose his experiment, even in case determinism and free will are incompatible. If t_0 is some time after the choice of the experiment, the experimenter has already made his choice before determinism is required to hold. Nevertheless, I wish I could do without determinism.

[3] See e.g. Earman (1986).

Now assume that the observer A is contained in the observed system O. Then every state of O uniquely determines a state of the subsystem A. To describe this I use a map R from S_O^j to S_A^j, where j is 0 or 1. In classical mechanics R is defined by discarding for each phase space point of O the coordinates referring to degrees of freedom not belonging to A; for probability distributions on phase space R is defined by taking the marginals over these degrees of freedom. In quantum mechanics R is the partial trace. (In general the partial trace of a pure state will not be any more a pure state but a density matrix. Therefore one should take as S_A^j and S_O^j the density matrices and not the vectors of the Hilbert space.) By putting $R(X) := \bigcup_{s \in X} R(s)$ the map R can be extended to a map $P(S_O^j) \to P(S_A^j)$, denoted also by R.

3 A counterexample

Given determinism and my (1995) result that it is impossible to measure from inside all *present* states exactly, one might be tempted to consider as trivial the result that one cannot measure from inside all past states. After all, given determinism there is a map from present states to past states—hence if I cannot distinguish present states it is pretty unsurprising that I cannot distinguish past states.

But there are two reasons why this is not trivial. For the measurement of present states from inside a consistency condition for the inference map I has to be required which is crucial in the proof:

$$R(I(s_A)) = \{s_A\} \tag{2}$$

should hold for all s_A. If this condition is violated I describes the inference of paradoxical conclusions from the pointer reading. (For suppose there is some $s \in I(S_A)$ with $R(s) \neq s_A$. Then one would infer from A at t_1 being in s_A that O at $t_0 = t_1$ might be in s and A in $R(s) \neq s_A$. There is a contradiction between A at t_1 being in s_A and at $t_0 = t_1$ it being in $R(s) \neq s_A$.) But (2) can only be imposed for the measurement of present states, where $t_0 = t_1$. If $t_0 \neq t_1$ it need not hold. Without (2) the proof in my (1995) fails.

The second reason why the result is not trivial is even better: a counterexample. Take as S_O^0 the pairs (n, i) where n is a natural number and i is either 0 or 1. Let $S_A^1 = S_A^0$ be the natural numbers, and the restriction map $R : (n, i) \mapsto n$. This represents for example a toy universe of two particles.

One particle, A, can be at positions $1, 2, \ldots$, the other could be a needle pointing up or down. As time evolution take $T : (n, i) \to (2n + i, 0)$.

This model has all essential features: the time evolution is deterministic; A is properly contained in O at t_0 because O has a degree of freedom which does not belong to A. Still, there is an inference map which can measure every state of O at t_0.

Take for example $I : S_A^1 \to S_O^0, m \mapsto (n, i)$ where n is the biggest natural number smaller or equal to $m/2$ and i is m modulo 2. Extend this to a map of the power sets by (1). This inference map can measure every state of O exactly: For every $(n, i) \in S_O^0$ there is some set $X_{(n,i)}$, namely $\{2n + i\}$, which refers uniquely to (n, i): $I(2n + i) = \{(n, i)\}$. I also can distinguish every two states $(n, i), (n', i')$ of S_O^0 which are different: (n, i) is in $I(2n + i)$ but not in $I(2n' + i')$, and (n', i') is in $I(2n' + i')$ but not in $I(2n + i)$.

4 The argument

Requirement (2) is acceptable only for the measurement of present states, but there is a different requirement appropriate for measurements of past states from inside.

Lemma 1 *Every consistent inference map I fulfils*

$$R(T(I(s_A))) = \{s_A\} \tag{3}$$

for all apparatus states $s_A \in S_A^1$.

Proof: If (3) were violated, there would be a state $s \in I(s_A)$ with $R(T(s)) \neq \{s_A\}$, and we would arrive at a contradiction in the following way. Assume that after the experiment, at t_1, we know that A is in s_A. Then according to the definition of I we conclude that O at t_0 was in some state in $I(s_A)$, possibly in s. But this is impossible because then A at t_1 would be in the state $R(T(s)) \neq s_A$. QED.

A map $P(S_A^1) \to P(S_O^0)$ not fulfilling (3) cannot be used as inference map because it would describe the inference of paradoxical conclusions from the pointer reading. In the special case that A wants to measure the *present* state of O we have $t_0 = t_1$ and T is the identity map. Then the consistency condition reduces to (2). If $t_0 = t_1$ my assumption of determinism becomes vacuous. So the argument against the measurability of present states from inside holds for stochastic and deterministic time evolutions.

Now I want to spell out the condition that A is a proper subsystem of O:

There are states s, s' of O at t_1 whose restriction to A coincides.

This I call the *assumption of the inside observer.* It requires that there are states $s \neq s' \in S_O^1$ for which $R(s) = R(s')$.

Lemma 2 *The assumption of the inside observer and determinism imply that there are states s_1, s_2 of O at t_0, $s_1 \neq s_2$, for which*

$$R(T(s_1)) = R(T(s_2)). \tag{4}$$

PROOF: Define $T^{-1} : S_O^1 \to P(S_O^0)$ by $s \mapsto \{r \in S_O^0 : T(r) = s\}$. From the assumption of the inside observer we know that there are states $s, s' \in S_O^1$, $s \neq s'$, for which $R(s) = R(s')$. Since T is surjective, neither $T^{-1}(s)$ nor $T^{-1}(s')$ is empty. Furthermore, one can see that $T^{-1}(s) \not\subset T^{-1}(s')$. For if this were not the case $\{s\} = T(T^{-1}(s)) \subset T(T^{-1}(s')) = \{s'\}$ contradicting $s \neq s'$. Similarily it can be argued that $T^{-1}(s') \not\subset T^{-1}(s)$. From both it follows that there are states $s_1 \in T^{-1}(s), s_2 \in T^{-1}(s')$ with $s_1 \neq s_2$. $s_1 \in T^{-1}(s)$ and $s_2 \in T^{-1}(s')$ imply $R(T(s_1)) \in R(T(T^{-1}(s)))$ and $R(T(s_2)) \in R(T(T^{-1}(s')))$. But since $R(T(T^{-1}(s))) = \{R(s)\}$ we have $R(T(s_1)) = R(s)$, and also $R(T(s_2)) = R(s')$. From $R(s) = R(s')$ it follows that $R(T(s_1)) = R(T(s_2))$. QED.

Theorem 1 *If a system O evolves deterministically from t_0 to t_1, then an apparatus which is contained in O cannot measure exactly at time t_1 all states of O at t_0.*

PROOF: To prove this indirectly assume that A at t_1 can measure all states of O at t_0 exactly: $(\forall s \in S_O)(\exists X \in P(S_A^1)) : I(X) = \{s\}$. This assumption will lead to a contradiction.

From Lemma 2 we know that there are object states $s_1 \neq s_2$ at t_0 with $R(T(s_1)) = R(T(s_2))$. If all states are exactly measurable there are $X, X' \in P(S_A^1)$ with $I(X) = \{s_1\}, I(X') = \{s_2\}$. Because of (1) there is an apparatus state $s_A \in X$ with $I(s_A) = \{s_1\}$ and another $s'_A \in X'$ with $I(s'_A) = \{s_2\}$. By repeated application of (3) we obtain

$$\begin{aligned}\{s_1\} = I(s_A) = I(R(T(I(s_A)))) = I(R(T(s_1))) = \\ = I(R(T(s_2))) = I(R(T(I(s'_A))) = I(s'_A) = \{s_2\},\end{aligned}$$

which contradicts $s_1 \neq s_2$. QED.

Lemma 3 *(3) implies that for all $s_A \in S_A^1$ we have: $I(s_A) = \{s \in S_O^0 : s \in I(S_A^1), R(T(s)) = s_A\}$.*

PROOF: Let $s \in I(s_A)$. (3) implies that $R(T(s)) = s_A$. Because of (1) from $s \in I(s_A)$ we infer that $s \in I(S_A^1)$. Therefore $I(s_A) \subset \{s \in S_O^0 : s \in I(S_A^1), R(T(s)) = s_A\}$.

On the other hand, take a $s \in S_O^0$ with $s \in I(S_A^1)$ and $R(T(s)) = s_A$. We have to show that $s \in I(s_A)$. Because of (1) from $s \in I(S_A^1)$ we infer that there is a $s'_A \in S_A$ with $s \in I(s'_A)$. Therefore $R(T(s)) \in R(T(I(s'_A)))$. By (3) this implies that $s_A = R(T(s)) \in R(T(I(s'_A))) = \{s'_A\}$. Thus $s_A = s'_A$ and $s \in I(s_A)$. QED.

Theorem 2 *If a system O evolves deterministically from t_0 to t_1, then an apparatus A which is contained in O cannot at time t_1 distinguish states s_1, s_2 of O at t_0 for which $R(T(s_1)) = R(T(s_2))$. (That there are such states is implied by Lemma 2.)*

PROOF: Assume that there were an inference map I and sets X_1, X_2 of apparatus states at t_1 with $I(X_1) \ni s_1 \notin I(X_2)$ and $I(X_2) \ni s_2 \notin I(X_1)$. This assumption will lead to a contradiction.

Because of (1) from $s_1 \in I(X_1)$ we infer that there is an apparatus state $s_A \in X_1$ at t_1 for which $s_1 \in I(s_A)$. Then (3) implies $R(T(s_1)) = s_A$. Furthermore, by (1), from $s_2 \notin I(X_1)$ it follows that for all $s \in X_1$ we have $s_2 \notin I(s)$. Since on the other hand $s_2 \in I(X_2)$ we have $s_2 \in I(S_A^1)$ and $R(T(s_2)) = R(T(s_1)) = s_A$. By Lemma 3 this implies that $s_2 \in I(s_A)$. This contradicts the fact that $s_2 \notin I(s)$ for all $s \in X_1$. QED.

5 Discussion

Why did I have to make the assumption of determinism? It would have been possible to define T not as a map $S_O^0 \to S_O^1$ but more generally as a two place relation on $S_O^0 \times S_O^1$. This would have been appropriate also for stochastic time evolutions.

But for stochastic time evolutions Lemma 1 does not hold. (3) can be violated without a contradiction arising. For example, if there is a $s \in I(s_A)$ with $R(T(s)) \ni s'_A \neq s_A$, then (3) is violated but a contradiction does not necessarily arise. Admittedly, s could have evolved into a state whose restriction to the apparatus is not s_A, but this does not contradict $s \in I(s_A)$ because it could have happened—and actually it did—that s evolves

into a state whose restriction to the apparatus in fact is s_A. Therefore $s \in I(s_A)$ does not contradict $R(T(s)) \ni s'_A \neq s_A$. Neither does Lemma 2 hold for stochastic time evolutions. The proof of Lemma 2 breaks down because $T(T^{-1}(s))$ just contains s but is not equal to $\{s\}$ for stochastic time evolutions. Although the proof presented here makes essential use of the assumption of determinism, an appropriate reformulation of Theorems 1 and **??** also holds for stochastic time evolutions.

Now what about the counterexample? Which of the assumptions of Theorems 1 and 2 is violated? It is the assumption of the inside observer which only holds at t_0 but not at t_1. Therefore the consequent of Lemma 2 fails: there are no two different states s_1, s_2 of O at t_0 for which $R(T(s_1)) = R(T(s_2))$.

Theorems 1 and 2 depend crucially on requiring the assumption of the inside observer to hold at t_1. If it holds just at t_0 the consequent of the theorems may fail, as the counterexample shows. But there are natural conditions under which the assumption of the inside observer holds either at both times or at none: the number of degrees of freedom of O (i.e. for a classical system the dimension of the phase space) should be conserved; or: the laws of motion should be time translation invariant. Both conditions are violated in the counterexample, where the second degree of freedom is frozen between t_0 and t_1.

References

Breuer T. (1995), "The Impossibility of Accurate State-Self-Measurements", *Philosophy of Science* **62**, 197-214

Breuer T. (1996), "Subjective Decoherence in Quantum Measurements", *Synthese* **107**, 1-17

Dalla Chiara M.L. (1977), "Logical Self-reference, Set theoretical paradoxes and the measurement problem", *Journal of Philosophical Logic 6:* 331

Earman J. (1986), *A Primer on Determinism*, Reidel: North Holland

Kampis Gy., Weibel P. (eds.) (1993), *Endophysics: The World From Within. A New Approach to the Observer-Problem with Applications in Physics, Biology, and Mathematics*, Santa Cruz: Aerial

Mittelstaedt P. (1993), "Measurement induced interrelations between quantum theory and its interpretation", pp. 269-280 in P. Busch, P. Mittel-

staedt, P. Lahti: *Proceedings of the Symposium on the Foundations of Modern Physics 1993*, Singapore: World Scientific

Penrose R. (1989), chaps. 9, 10, *The Emperor's New Mind*, Oxford: University Press

Peres A. and Zurek W.H. (1982), "Is quantum mechanics universally valid?", *American Journal of Physics* **50**, 807

Popper K.R. (1950), "Indeterminism in classical physics and quantum physics", *British Journal for the Philosophy of Science* **1**, 173

Primas H. (1990), "Mathematical and Philosophical Questions in the Theory of Open Quantum Systems", pp. 233-258, in A.I.Miller (ed.): *Sixty-two Years of Uncertainty:Historical, Philosophical and Physics inquiries into the Foundations of Quantum Mechanics*, New York: Plenum

Rössler O.E. (1987), "Endophysics", in J.L.Casti and A.Karlqvist (eds.): *Real Brains—Artificial Minds*, New York: North-Holland

Svozil K. (1993), *Randomness and Undecidability in Physics*, Singapore: World Scientific