PHYSICAL REVIEW D VOLUME 44, NUMBER 10 15 NOVEMBER 1991

Quantum mechanics near closed timelike lines

David Deutsch
Oxford University Mathematical Institute, 24-29 St. Giles, Oxford OX1 3LB, England
(Received 9 April 1991)

The methods of the quantum theory of computation are used to analyze the physics of closed timelike lines. This is dominated, even at the macroscopic level, by quantum mechanics. In classical physics the existence of such lines in a spacetime imposes "paradoxical" constraints on the state of matter in their past and also provides means for knowledge to be created in ways that conflict with the principles of the philosophy of science. In quantum mechanics the first of these pathologies does not occur. The second is mitigated, and may be avoidable without such spacetimes being ruled out. Several novel and distinctive (but nonparadoxical) quantum-mechanical effects occur on and near closed timelike lines, including violations of the correspondence principle and of unitarity. It becomes possible to "clone" quantum systems and to measure the state of a quantum system. A new experimental test of the Everett interpretation against all others becomes possible. Consideration of these and other effects sheds light on the nature of quantum mechanics.

APPLYING THE QUANTUM THEORY OF COMPUTATION

This paper is about the physical effects of closed timelike lines. Traditional discussions (e.g., Refs. [1–5]) treat such effects as properties of spacetime geometry and use the methods of differential geometry and general relativity. There are two main deficiencies in that approach. First, one of the principal theoretical problems concerning closed timelike lines is that of distinguishing merely counterintuitive effects from downright unphysical ones, but chronology-violating spacetimes (i.e., spacetimes containing closed timelike lines) tend also to have other unfamiliar features such as "wormholes" and singularities. These may or may not persist when quantum gravity is taken into account. And they introduce technical and conceptual problems of their own which make it difficult to be confident that a particular spacetime model correctly represents the effects that chronology violation would have if it actually occurred.

The second, more profound deficiency is that classical spacetime models do not take account of quantum mechanics which, even aside from any effects of quantum gravity, actually dominates both microscopic and macroscopic physics on and near all closed timelike lines.

My approach will be through the quantum theory of computation (see Refs. [6,7]). I abstract away most of the underlying geometry and consider only the world lines of finitely many particles. I neglect the dynamics of the motion of the particles, approximating that as classical and given. Only the internal degrees of freedom of the particles are treated quantum mechanically. These interact only during periods when the world lines of the particles are very close together in space. The resulting model, that of a finite set of finite-state systems traveling along fixed trajectories and interacting only at short range, may seem restrictive and artificial, but the following two considerations should counteract that impression. First, the class of such models is essentially the class of *quantum computational networks* [7], which is computationally universal in the sense that such networks can simulate the behavior of any finite quantum system. Allowing the networks to have temporal as well as spatial loops can be expected to extend the universality in the appropriate way. Second, for any quantum computational network there is a spacetime such that the trajectories of particles through the network can be identified with a set of correspondingly connected and aligned timelike world-line segments of the spacetime. For example, the trajectory of a particle through the chronology-violating network shown in Fig. 1(a) can be identified with the looping timelike world line in the spacetime region shown in Fig. 1(b).

In a spatial diagram of a quantum computational network, such as Fig. 1(a), chronology-violating links introduce a negative delay time, as indicated by the circled "-1." Negative delay components in the model play the role of *time machines,* which I define in general as objects in which some phenomenon characteristic only of chronology violation can reliably be observed. (To capture the intuitive notion of a time machine one would add the requirement that the phenomenon be "macroscopic.")

One must be careful to distinguish purely spatial diagrams such as Fig. 1(a) from spacetime diagrams such as Fig. 1(b). Figure 2 shows how two trivial quantum computational networks, namely a spatial loop and a closed timelike loop, are represented in spatial and spacetime diagrams.

The basic method of this paper is to regard computations as representative physical processes—representing the behavior of general physical systems under the unfamiliar circumstances of chronology violation. Computations are usually performed with the intention of creating an *output* that has certain desired properties depending on an *input.* But we are interested in the physical

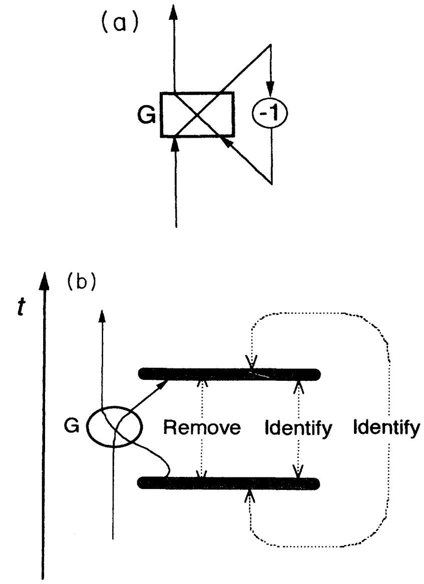

FIG. 1. (a) Chronology-violating network with gate G and negative delay. (b) Corresponding chronology-violating spacetime with interaction region G.

evolution of computers, not primarily in what they are computing. To the user the inputs and outputs refer to something in a problem domain, so in order to use a computing machine the user attaches meanings from the problem domain to states in which the machine can be prepared and measured. To the user, some of the states through which the machine may evolve are distinguished as satisfactory output states, and the machine must signal if and when it has reached such a state, because, in general, a computation has an unpredictable duration. It is then said to have "halted." In the theory of computation one is principally interested in various properties of halting and halted states. But physically the halted state has no special significance, so for our present purpose it is convenient to treat halted states uniformly with others. I shall therefore consider networks bounded not only in space but also by initial and final spacelike hypersurfaces, and take the input and output of a network to be its states on the initial and final hypersurfaces, respectively.

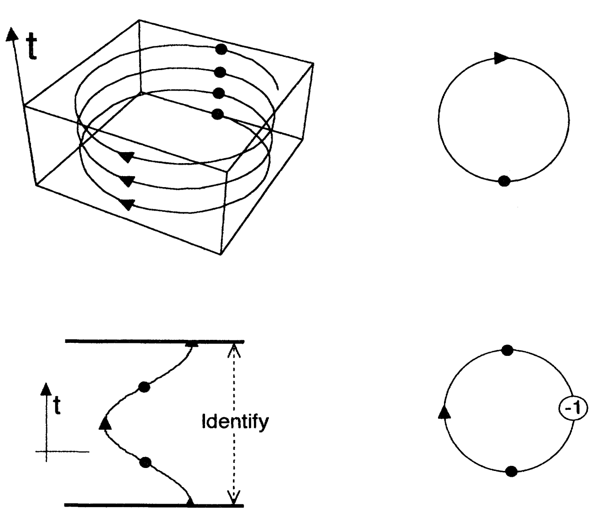

FIG. 2. Spatial and spacetime representations of computational networks: Spacetime representation of a spatial loop (top left), spatial representation of that loop (top right), spacetime representation of a closed timelike loop (bottom left), spatial representation of that loop (bottom right).

The *chronology-violating region* of a spacetime is the set of events through which closed timelike lines pass. The remainder of the spactime is its *chronology-respecting region.* The methods of this paper are applicable whether or not there is a chronology-respecting region, but I shall proceed initially on the assumption that there is, and furthermore that each connected chronology-violating region has an *unambiguous past* and an *unambiguous future* which are the chronology-respecting portions of the chronological past and future, respectively of that region. Later I shall comment on spacetimes that violate these assumptions.

Let us call two spacetime-bounded networks *denotationally equivalent* if their output states are the same function of their input states (under a give one-to-one correspondence between their states) even if they do not produce the outputs from the inputs in the same way, or in the same time. It sometimes happens when chronology is violated that the output is not determined uniquely by the input but depends on supplementary data in the chronology-violating region. Later I shall suggest a way of fixing such ambiguities, but in the meantime let us call two networks denotationally equivalent if for each possible input state the *sets* of possible output states, as the supplementary data range over all possible values, are the same.

A *denotationally trivial transformation* of a network is one that results in a network that is denotationally equivalent to the original one. The members of a denotational equivalence class respond equivalently to stimuli in that they create equivalent relationships between outputs and inputs. Insofar as we wish to study chronology-violating regions solely in terms of their external responses to external stimuli (the equivalent of S-matrix theory) we are entirely free to use denotationally trivial transformations to simplify networks that model the effects of the region. But it must be borne in mind that as regards what happens *inside* a chronology-violating region, a computational network that faithfully models the physics of the region may cease to do so under certain denotationally trivial transformations. Therefore when transforming networks one should where appropriate take note of what the transformations correspond to in the spacetime that is being modeled.

General spacetime-bounded networks may be converted by denotationally trivial transformations into a simplified standard form as follows. Each particle traveling in the network is replaced by sufficiently many particles each of which carries one 2-state internal degree of freedom, or *bit.* The particle itself is called the "carrier" of the bit. The nonoverlapping regions in which bits interact are called *gates*. There are denotationally trivial transformations that localize all self-interactions into gates, so that bits are inert when traveling between gates.

Chronology violation in itself (i.e., the existence of negative delays) makes no fundamental difference to the behavior of a network unless there is a *closed path for information*. Such a path is not necessarily the trajectory of any carrier in the network because bits on different carriers interact and can exchange information in gates. If a network has no closed path for information there is always a denotationally trivial transformation that removes all negative delays by introducing positive delays elsewhere. If the network does have such a path then there is always a denotationally trivial transformation that makes all negative delays occur on *closed* trajectories of carriers.

Consider any gate of the network, other than sources and sinks, which we may as well consider to be additional inputs and outputs. Since the gate is reversible the same number of carriers emerge from it as enter it. For each carrier whose trajectory is closed, there is a denotationally trivial transformation that replaces each of the negative delays occurring on the trajectory by a trajectory that goes first into the (ambiguous) future of all gates of the network then with a single large negative delay back into the (ambiguous) past of all the gates, and then forward to wherever and whenever it was going. All negative-delay components then consist of trajectories, uninterrupted by gates, from the future to the past of all gates, and the network has the following character: A group of m bits (the input) enters the chronology-violating region from the unambiguous past, and interacts with a group of n bits whose carriers are on closed timelike trajectories. After the interaction the m bits continue into the unambiguous future, forming the output. Further denotationally trivial transformations rearrange the timings so that all the negative delays are the same, say $-T$, where T is the time required for the gates, considered as a single gate G, to perform their computation. The network then has the form shown in Fig. 3 (the meanings of $\hat{\rho}_1$ and $\hat{\rho}_2$ will be given below).

Sometimes it is more illuminating to stop short of transforming the network fully into the form of Fig. 3, and to allow some of the input and output carriers to be among those that travel back in time. For example, the fully transformed version of the network in Fig. 1(a) would contain two carriers instead of one, with one of them on a closed timelike trajectory. With a suitably transformed gate, that would indeed be denotationally

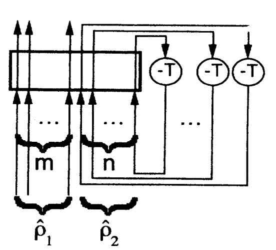

FIG. 3. General spacetime-bounded chronology-violating network after the transformations described in the text.

equivalent to Fig. 1(a), and either version of the network would serve in an analysis of the other's externally observable behavior. But the transformed version would be intuitively very different from the original one which might represent a time traveler, whereas the transformed version appears to represent an ordinary space traveler meeting a time traveler who spontaneously comes into existence as an identical twin of the space traveler, exists for a finite period on an "eternal" loop, and then ceases to exist.

When the network is in the form of Fig. 3 we have separated its elements into two regions. On the left is an interaction region in which chronology is "locally" respected (i.e., there are no negative delays), and on the right is a region of chronology-violating links in which all bits are inert and all carriers are on trajectories closed in spacetime. All the special properties of chronology-violating networks are consequences of consistency conditions around those closed trajectories.

CLASSICAL PARADOXES OF CHRONOLOGY VIOLATION

Intuitive arguments against chronology violation

Chronology-violating spacetimes are customarily deemed to be unphysical, even if (like the Gödel universe or the maximal Kerr solution) they obey the Einstein equations. The conventional reason is given, for example, by Hawking and Ellis [3] who argue that "...the existence of [closed timelike] curves would seem to lead to the possibility of logical paradoxes: for, one could imagine that with a suitable rocketship one could travel round such a curve and, arriving back before one's departure, one could prevent oneself from setting out in the first place. Of course, there is a contradiction only if one assumes a simple notion of free will; but this is not something which can be dropped lightly since the whole of our philosophy of science is based on the assumption that one is free to perform any experiment."

There are two distinct arguments here. The first attempts to rule out chronology-violating spacetimes on logical grounds alone, which is a fallacy (cf. Ref. [8]). The second begins by pointing out that fallacy. What Hawking and Ellis are getting at in the second argument is that classically there are more constraints on initial data in chronology-violating spacetimes than in chronology-respecting ones. We shall see that if there are closed timelike lines to the future of a given spacelike hypersurface, the set of possible initial data for classical matter on that hypersurface can be heavily constrained compared with what it would be if the same hypersurface with the same local interactions were embedded in a chronology-respecting spacetime.

This may not be what we are used to, but can we really know *a priori* that such effects do not occur in nature? As I shall explain, in classical physics chronology violations would indeed make a spacetime unphysical because they would conflict with "our whole philosophy of science," but neither argument of Hawking and Ellis correctly explains why. Their appeal to free will, if it were valid, would rule out not only chronology-violating spacetimes but all spacetimes that obey classical laws of motion such as the Einstein equations—yet those are not usually deemed to violate philosophical principles. For one might as well say that the standard Robertson-Walker spacetime conflicts with our whole philosophy of science because its inhabitants too are not "free to perform any experiment" but are obliged to perform precisely those experiments that the initial conditions dictate.

In any case, retrospective constraints placed on initial data by chronology violations are a peculiarity of classical physics. I shall show that there is no analogous constraint in quantum mechanics.

Classical analysis of the paradoxes

In order to introduce the method and notation, let us consider the classical case first. But I must warn the reader at the outset that the arguments and conclusions of this section are based on the premise that classical physics is at least approximately true near closed timelike lines, and we shall find that that premise is false.

A classical computational network is one whose gates all effect Turing operations so that outside the gates the state of each bit is always an element of a certain fixed basis, the computational basis. We may also assume that all the gates are reversible, since no significant generality is lost by considering only reversible computations [9].

Paradox 1

Suppose that a single carrier travels along the world line in Fig. 1(b) with its bit initially in a state $|x\rangle$ of the computational basis, where $x \in \mathbb{Z}_2$ (the set $\{0,1\}$ of integers modulo 2). Suppose also that the gate G is a *measurement gate*, which causes the two versions of the bit passing through it to undergo the interaction

$$|x\rangle|y\rangle \Longrightarrow |x \dot{+} y\rangle|y\rangle \quad (\forall x,y \in \mathbb{Z}_2) \tag{1}$$

with each other. The first and second kets in the representation refer to the younger and older versions of the bit, respectively, with respect to the proper time of the carrier. The "evolves into" symbol "$\Longrightarrow$" means that if the bits enter the gate in the state to the left of the symbol, they leave it in the state to the right. "$\dot{+}$" is the exclusive-or operation, or bit-by-bit addition modulo 2.

One possibility is plainly that the bit has the value $x=0$ initially and retains that value throughout. But can the bit have the initial value $x=1$? Intuitively the answer is no, for if it did, in what state $|y\rangle$ would it first (in terms of its proper time) emerge from the gate? The interaction (1) would ensure that that state was different from the one in which it then (in terms of its proper time) entered, but that is a contradiction because under the circumstances those two states are alternative ways of specifying the same thing.

More precisely, one can read off from (1) that on leaving the gate G in Fig. 1(a) the younger version of the bit is in the state $|x \dot{+} y\rangle$ and the older version is still in the state $|y\rangle$ in which it entered. But the kinematics of the situation require that the younger version leave the gate in the same state in which the older one enters, for the bit does not evolve outside the gate. Therefore

$$x \dot{+} y = y \ . \tag{2}$$

This confirms the intuitive argument that the chronology violation retrospectively places a constraint on the initial value x of the bit. In this case the initial value must be zero, for no value of y satisfies the consistency condition (2) if $x=1$.

Paradox 2

A drastic constraint is imposed if the gate G in Fig. 1(a) effects the evolution

$$|x\rangle|y\rangle \Longrightarrow |y \dot{+} 1\rangle|x\rangle \quad (\forall x,y \in \mathbb{Z}_2) \ . \tag{3}$$

The kinematical consistency condition is

$$y \dot{+} 1 = y \ , \tag{4}$$

which cannot be satisfied and therefore rules out all initial values x.

This corresponds to the nonoccurrence of the following apparently paradoxical history: while the younger and older versions of a time traveler are interacting, the younger one makes a note of something that the older one says, and then, after traveling back in time, deliberately fails to say that thing. This proves that according to classical physics people (or automata) who would behave in the way just described when traveling in time do not travel in time.

In the terms of our model, the fact that (4) is a contradiction implies that the bit does not enter the gate, so our decision to treat the motion of the bit's carrier as given turns out to be unsustainable if the interaction is (3).

Paradox 3

Particles whose trajectories are changed by forces acting on them can be simulated in a computational network model by using an "occupation number" representation. Consider two-possible trajectories of a particle through the chronology-violating region, the right-hand one going back in time and the left-hand one merely passing by. This is simulated by the network shown in Fig. 4. The

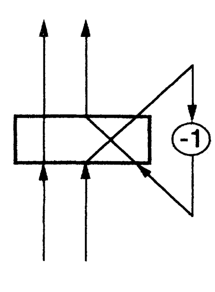

FIG. 4. Network for the traditional paradox of time travel.

network is traversed by two carriers. The first carrier takes the left-hand, and the second the right-hand trajectory. The bit on each carrier is 1 or 0 according to whether the particle being simulated would be present on that carrier's trajectory or not. Now we can treat the traditional paradox of a particle which, in the words of Hawking and Ellis "travels round such a curve and, arriving back before its departure . . . prevents itself from setting out in the first place." The two bits in the network are initially in the state $|0\rangle|1\rangle$, simulating a system in which the left-hand trajectory is unoccupied and the right-hand trajectory, which loops, is occupied by a particle. Before going their separate ways, the two carriers encounter an older version of the right-hand one, which acts on them according to

$$|x\rangle|x\dot{+}1\rangle|y\rangle \Longrightarrow |x\dot{+}y\rangle|x\dot{+}y\dot{+}1\rangle|y\rangle \quad (\forall x,y\in\mathbb{Z}_2) \ . \tag{5}$$

In terms of the simulated system this says that if the younger version of the particle meets the older version it changes course from either of its two possible trajectories to the other, but if it does not meet the older version it does not change course. This time the kinematical consistency condition is

$$x\dot{+}y\dot{+}1=y \ , \tag{6}$$

which implies that $x=1$; i.e., the particle must start on the left trajectory and cannot go back in time.

Under this classical analysis the logic of paradox 3 is the same as that of paradox 1. Because we are using the network of Fig. 4 to simulate one particle with two possible trajectories, the initial and final states of the network must be either $|0\rangle|1\rangle$ or $|1\rangle|0\rangle$. The other two possibilities do not simulate any state of the particle. Under the correspondence

$$|x\rangle|x\dot{+}1\rangle \sim |x\rangle \ , \tag{7}$$

interaction (5) becomes interaction (1), and condition (6) becomes condition (2).

Paradox 4 and the evolutionary principle

This paradox is similar to paradox 1, but with the *consistent* initial value $x=0$. To understand why the consistent initial value is paradoxical—potentially much more so than the inconsistent value—consider first the experiment being performed with a large number n of carriers simultaneously, with each bit having the initial value $x=0$ and interacting according to (1) with a future version of itself. A consistency condition of the form (2) then holds for each bit, and each condition has two solutions, namely $y=0$ and $y=1$. Therefore even though initial data have been completely specified in the unambiguous past, the n chronology-violating bits can still jointly be in any of their 2^n possible states, and the same is true of the n output bits.

Thus in classical physics a chronology-violating system is in general both overdetermined and underdetermined by the usual initial data—overdetermined because certain initial data which would otherwise be permitted are forbidden, and underdetermined because additional data about what happens in the chronology-violating region may be required to specify uniquely a solution of the dynamical equations. The term "initial data" is a misnomer for data which cannot be set on the past boundary of spacetime, so when I wish to draw attention to their noninitial character I shall call them *supplementary data.*

Some of the choices of supplementary data are apparently unexceptionable; for example, as I have said, if $y=0$ for each bit all the bits pass through the entire process unchanged. On the other hand, it is philosophically unacceptable for the bits to be in certain highly complex states, specifically states which encode the solutions of difficult problems. The difficulty is illustrated by the following history: A time traveler goes into the past and reveals the proof of an important theorem to the mathematician who had later been recognized as the first to prove it. The mathematician goes on to publish the proof, which is then read by the time traveler before setting out. Who thought of the proof? No one, since each of the two participants obtained that valuable information from the other.

It is a fundamental principle of the philosophy of science [10,11] that the solutions of problems do not spring fully formed into the Universe, i.e., as initial data, but emerge only through evolutionary [12] or rational processes. In adopting this *evolutionary principle* we reject such antirational doctrines as creationism, and more generally we reject all explanations of complex regularity in the observed Universe that attribute it to complex regularity in the initial data.

That one rules out certain logically and empirically possible initial data for philosophical reasons is in no way paradoxical or improper. We do precisely that for the initial data at the big bang, even though any number of cosmologies consistent with observation but violating the evolutionary principle can be constructed [13]. Attempts have been made to codify the conjecture that the initial data for the Universe are very simple (e.g., Ref. [14]), and similarly for naked singularities [15]. These may be seen in part as attempts to implement the evolutionary principle.

Paradox 4 is that in the presence of chronology violations the evolutionary principle can conflict with the requirement that the evolution be kinematically consistent. Consider the $2n$-bit gate G_f in the network shown in Fig. 5. The first group of n input bits of G_f, represented by the line entering G_f on the left, specifies an argument $x\in\mathbb{Z}_{2^n}$ for a function f. The action of G_f is to replace the last n bits of its input, represented by the line entering G_f on the right, by the exclusive-or of those bits with $f(x)\in\mathbb{Z}_{2^n}$. In other words G_f has the effect

$$|x\rangle|y\rangle \Longrightarrow |x\rangle|y\dot{+}f(x)\rangle \quad (\forall x,y\in\mathbb{Z}_{2^n}) \ . \tag{8}$$

If $P\neq NP$ (see Ref. [16]) the task of finding a fixed point of f, i.e., a value x' such that $f(x')=x'$, is, in general, much harder (requires significantly more computational resources) than computing one value $f(x)$. The evolutionary principle therefore requires that no network that uses few computational resources other than a single run-

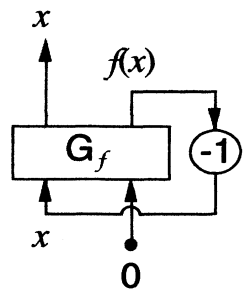

FIG. 5. Network for finding a fixed point of f?

ning of gate G_f should be able reliably to produce the output x' starting with n zero bits as input.

But that is precisely what the network of Fig. 5 seems to do. If y is prepared with the initial value 0 the kinematical consistency condition is simply that $x=f(x)$, so that the n output bits of the network must jointly have the value of a fixed point x' of f.

What has happened here is that the kinematical consistency condition is "almost" inconsistent. If f has no fixed point it is strictly inconsistent and we have another version of paradox 2, in which case the initial data must conspire to prevent the experiment from being performed at all, an outcome which could itself be detected and interpreted as meaning that f has no fixed point. If f is the identity function so that every value $x \in \mathbb{Z}_{2^n}$ is a fixed point, we have the apparently unproblematic consistent version of paradox 1, as just discussed. But if f has exactly one fixed point, the supplementary data in the chronology-violating region have no option but tamely to set themselves to the knowledge-laden value x' ready to feed themselves self-consistently into the gate.

This "near inconsistency," forcing a violation of the evolutionary principle, is a far more serious paradox than the "actual" inconsistencies of paradoxes 1–3. Those inconsistencies merely indicate that the initial data have one set of values rather than another, something which is true anyway, and starting from those values the subsequent evolution, though strange, does not contradict the philosophy of science. But because of the "near inconsistency" of paradox 4 the only permitted initial data cause an evolution that does contradict the philosophy of science.

All the above conclusions about what would happen under the conditions of paradoxes 1–4 are more or less what one would expect from intuitive classical reasoning. That is why nearly all of them are false. Now I shall analyze the same situations correctly, using quantum mechanics.

QUANTUM COMPUTATIONAL ANALYSIS OF THE PARADOXES

Paradox 1

Let us reconsider paradox 1 without making the unwarranted assumption that the two versions of the particle emerge from the interaction with their bits in a classical (computational basis) state. Let the initial state of the bit be $|\psi\rangle$, which may or may not be an element of the computational basis. After the bit has interacted with the older version of itself each version individually may be in a mixed state [17]. Let $\hat{\rho}$ be the density operator of the older bit as it enters the gate. Then the joint density operator of the two bits entering the gate is

$$|\psi\rangle\langle\psi|\otimes\hat{\rho} \ . \tag{9}$$

Because the evolutions undergone by the bits in paradoxes 1–4 are locally reversible, the action of the gate in each case is described by a unitary operator U on the joint state space. In the case of paradox 1 we can read off from (1) that

$$U=\sum_{x,y\in\mathbb{Z}_2} |x \dot{+} y\rangle|y\rangle\langle x|\langle y| \ . \tag{10}$$

The density operator of the two bits as they leave the gate is

$$U(|\psi\rangle\langle\psi|\otimes\hat{\rho})U^{\dagger} \ . \tag{11}$$

Kinematical consistency now implies that the density operator of the younger bit as it leaves the gate is the same as that of the older bit as it enters the gate, for those bits are one and the same, i.e.,

$$\mathrm{Tr}_2[U(|\psi\rangle\langle\psi|\otimes\hat{\rho})U^{\dagger}]=\hat{\rho} \ , \tag{12}$$

where the trace is over the subspace of states of the older bit.

Unlike its classical analogue (2), the condition (12) has solutions for all initial states $|\psi\rangle$, including the classically forbidden one $|\psi\rangle=|1\rangle$. A little algebra shows that if U is given by (10),

$$\hat{\rho}=\tfrac{1}{2}\hat{1}+\mathrm{Re}(\langle 0|\psi\rangle\langle\psi|1\rangle)(|0\rangle\langle 1|+1\rangle\langle 0|) \tag{13}$$

is a solution of (12) for general $|\psi\rangle$. For $|\psi\rangle=|0\rangle$ a family of additional solutions exists, which I shall discuss below. From (10), (11), and (13) we can calculate the final density operator of the particle, namely

$$\mathrm{Tr}_1[U(|\psi\rangle\langle\psi|\otimes\hat{\rho})U^{\dagger}] =\tfrac{1}{2}\hat{1}+2[\mathrm{Re}(\langle 0|\psi\rangle\langle\psi|1\rangle)]^2(|0\rangle\langle 1|+|1\rangle\langle 0|) \ . \tag{14}$$

Thus the chronology violation has placed no constraint on the initial state $|\psi\rangle$.

I shall now sketch a proof showing that there is no constraint on quantum initial data not only for the interaction just discussed, but for general interactions U.

Consider a general spacetime-bounded network after simplification by denotationally trivial transformations, as shown in Fig. 3. Let the initial density operator of the m non-time-traveling bits be $\hat{\rho}_1$, an operator on the 2^m-dimensional state space $\mathcal{H}_1$ of those bits, and let the density operator of the n time-traveling bits be $\hat{\rho}_2$, an operator on the 2^n-dimensional state space $\mathcal{H}_2$ of those bits. The kinematical consistency condition is that the latter n bits enter the gate with the same density operator with which they left it, i.e.,

$$\mathrm{Tr}_1[U(\hat{\rho}_1 \otimes \hat{\rho}_2)U^\dagger] = \hat{\rho}_2 \ . \tag{15}$$

The initial state $\hat{\rho}_1$ will be unconstrained if and only if for every density operator $\hat{\rho}_1$ there is a solution $\hat{\rho}_2$ of (15). The expression on the left in (15) may be regarded as the image of $\hat{\rho}_2$ under a linear superscattering operator S on the space of density operators on $\mathcal{H}_2$, defined by

$$\mathsf{S}* = \mathrm{Tr}_1[U(\hat{\rho}_1 \otimes *)U^\dagger] \ , \tag{16}$$

where "$*$" indicates the position of the operand of the superscattering operator, so it remains to be proved that every operator of the form (16) has a fixed point.

Let $\hat{\rho}(0)$ be any density operator on $\mathcal{H}_2$, and for each integer N define

$$\hat{\rho}(N) \equiv \frac{1}{N+1} \sum_{n=0}^{N} \mathsf{S}^n \hat{\rho}(0) \ . \tag{17}$$

Each $\hat{\rho}(N)$ is the mean of $N+1$ density operators, and is therefore itself a density operator on $\mathcal{H}_2$. Since

$$\begin{aligned} 0 \le E(\hat{\rho}(N)) &\equiv \mathrm{Tr}\{[\mathsf{S}\hat{\rho}(N) - \hat{\rho}(N)]^2\} \\ &= \frac{1}{(N+1)^2} \mathrm{Tr}[(\mathsf{S}^{N+1}\hat{\rho}(0) - \hat{\rho}(0))^2] \\ &\le \frac{1}{(N+1)^{2'}} \end{aligned} \tag{18}$$

the sequence $[E(\hat{\rho}(N))]$ has zero as its greatest lower bound. Because the space of density operators on a given finite-dimensional state space is compact, the sequence $\{\hat{\rho}(N)\}$ must have one or more accumulation points $\hat{\rho}_{\lim}$. For each one of these $E(\hat{\rho}_{\lim}) = 0$ by continuity from (18), so each $\hat{\rho}_{\lim}$ is a fixed point of the superscattering operator S. That completes the proof of the proposition that within the limitations of this model (which we have some reason to believe is universal) *closed timelike lines place no retrospective constraint on the state of a quantum system.*

Paradox 2

It can be shown by the same method that under the interaction (3) of paradox 2 the density operator $\hat{\rho}$ of the chronology-violating bit must have the form

$$\tfrac{1}{2}[\hat{1} + \lambda(|0\rangle\langle 1| + |1\rangle\langle 0|)] \tag{19}$$

for any real λ with $0 \le \lambda \le 1$. As λ varies (19) ranges from a pure state

$$\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \tag{20}$$

to the maximally mixed state $\frac{1}{2}\hat{1}$. The final state, regardless of λ, is equal to the initial state $|\psi\rangle$.

Paradox 3

The traditional paradox 3 is again similar to paradox 1 under the correspondence (7), but no longer quite equivalent. This time the pure input states that simulate a single particle on one of two trajectories are those in the subspace spanned by the states $|0\,1\rangle$ and $|1\,0\rangle$ [18] of the two input bits of the network.

However, it is no longer required that the final state be in that subspace. The output $|1\,1\rangle$ simulates a 2-particle state, with one particle on each trajectory, and the output $|0\,0\rangle$ simulates the 0-particle state. The process conserves the expectation value of the particle number operator

$$\hat{N} = |0\,1\rangle\langle 0\,1| + |1\,0\rangle\langle 1\,0| + 2|1\,1\rangle\langle 1\,1| \tag{21}$$

(we shall see later that this is no accident) but the final state need not be an eigenstate of $\hat{N}$ even though the initial state is. If the state of the two input bits is $|\psi\rangle$, the kinematical consistency condition for the density operator $\hat{\rho}$ of the chronology-violating bit is

$$\mathrm{Tr}_{1,3}[U(|\psi\rangle\langle\psi| \otimes \hat{\rho})U^\dagger] = \hat{\rho} \ , \tag{22}$$

where the trace is over the state spaces of the first bit and the older version of the second bit.

If $|\psi\rangle$ is the classically forbidden state $|0\,1\rangle$, the only solution of (22) is $\hat{\rho} = \frac{1}{2}\hat{1}$, and that makes the density operator of the two output bits

$$\mathrm{Tr}_2[U(|\psi\rangle\langle\psi| \otimes \hat{\rho})U^\dagger] = \tfrac{1}{2}(|0\,0\rangle\langle 0\,0| + |1\,1\rangle\langle 1\,1|) \ . \tag{23}$$

The probability that there will be exactly one particle in the output is zero, whereas in the classical case it was a certainty that there would be exactly one particle.

Paradox 4

In paradox 4 the kinematical consistency condition in the quantum case is no longer that the value represented by the chronology-violating bits be a fixed point x' of the function f but that the density operator $\hat{\rho}$ of those bits satisfy

$$\mathsf{S}_f \hat{\rho} = \hat{\rho} \ , \tag{24}$$

where S_f is the superscattering operator defined by

$$\mathsf{S}_f * = \sum_{x \in \mathbf{Z}_2{}^n} [(|f(x)\rangle\langle x|) * (|x\rangle\langle f(x)|)] \ . \tag{25}$$

The analogue $|x'\rangle\langle x'|$ of the classical solution is always a solution of (24), though for most functions f it is not the only one. Let us find the other solutions, in the hope that some of them may satisfy the evolutionary principle. By inspection of (25), every solution of (24) is diagonal in the computational basis, that is,

$$\hat{\rho} = \sum_{x \in \mathbf{Z}_2{}^n} p_x |x\rangle\langle x| \ , \tag{26}$$

where

$$\sum_{x \in \mathbf{Z}_{2^n}} p_x = 1 \quad \text{and} \quad 0 \le p_x \le 1 \ .$$

Substituting (26) and (25) into (24) and comparing coefficients we obtain

$$p_x = \sum_{\{y | x = f(y)\}} p_y \ . \tag{27}$$

Let us call an integer $x \in \mathbb{Z}_{2^n}$ "idempotent of degree k under f" and define

$$D_f(x)=k \tag{28}$$

if k is the least positive integer such that

$$f^{(k)}(x) \equiv \overbrace{f(f(\cdots f(x)))}^{k \text{ times}} = x \quad . \tag{29}$$

$D_f(x)=0$ if there is no such positive integer, and $f^{(0)}(x)$ is defined as x. Thus any fixed point x' of f has $D_f(x')=1$. For any f there is at least one idempotent value under f.

If f is invertible, (27) implies that

$$p_x = p_{f(x)} \quad (\forall x \in \mathbb{Z}_{2^n}) \tag{30}$$

so the kinematical consistency condition (24) amounts to the requirement that $\hat{\rho}$ be a sum of terms of the form

$$p_x \sum_{i=0}^{D_f(x)-1} |f^{(i)}(x)\rangle\langle f^{(i)}(x)| \quad . \tag{31}$$

The maximally mixed state $2^{-n}\hat{1}$ is such a sum; so is the classical solution $|x'\rangle\langle x'|$, and in general there are many others. Supplementary data are required. Before continuing the investigation of the solutions of (24) I shall propose a rule for fixing supplementary data in general.

The evolutionary principle and supplementary data

Any rule that specifies supplementary data (and indeed any rule that specifies initial data) must conform to the evolutionary principle. The evolutionary principle itself is therefore the natural starting point for attempts to find such a rule. Unfortunately there is at present no precise formulation of the evolutionary princple. An informal statement of it is:

Knowledge comes into existence only by evolutionary processes.

It is understood here that biological adaptive complexity and similar quantities are forms of "knowledge" and processes such as rational thought are deemed, insofar as they succeed in generating knowledge, to be in a generalized sense "evolutionary."

Applying this informal version to our present problem, we infer that the supplementary data in a chronology-violating region must contain no knowledge over and above what was in the initial data on a spacelike hypersurface immediately to the unambiguous past of the region. This is reminiscent of Penrose's "cosmic censorship" hypothesis: although we might be able to see into the interior of the chronology-violating region, we can see nothing "interesting" there unless it had already been there before or, perhaps, if it had had enough (proper) time to evolve within the region.

I am led to suggest the following *maximum entropy rule* [19] for supplementary data:

The state of the supplementary data (i.e., data required elsewhere than at the past boundary of spacetime for fixing a global solution of the dynamical equations) is the state of greatest entropy compatible with the initial data.

Knowledge is not the same thing as information, nor is it any function of information alone. There is as yet no quantitative measure of "knowledge" that could be incorporated into physics. However, it is reasonable to suppose that the requirement that a system contain no independent information (which is what the maximum entropy rule effectively says) might also imply that the system contains no independent knowledge.

Another satisfactory property of the maximum entropy rule is that it is independent of basis in the space of states. This property is not shared by current complexity-theoretic definitions of what I am calling "knowledge."

I conjecture that the maximum entropy rule has a further property, that of being "transparent" to denotationally trivial changes in a network. That is, if the rule is applied to fix the supplementary data for each member of a set of denotationally equivalent networks, they remain denotationally equivalent.

I have referred to the evolutionary principle as a "philosophical" principle. That is so, but note that any specific implementation of it that predicts the behavior of physical systems, as the maximum entropy rule does, is in principle a testable proposition whose status is no different from that of any other physical theory.

Let us see how the maximum entropy rule fixes the supplementary data in simple cases. First, note that in the case of a bit that loops in the manner of Fig. 1(a) but does not interact with its younger self, no supplementary data are required and the output is equal to the input—which is presumably the correct answer. After a denotationally trivial change, the network connects the output directly to the input, but there is a second bit on a disconnected closed timelike line. The maximum entropy rule then places that bit in the maximally mixed state $\frac{1}{2}\hat{1}$. The output is unchanged by the denotationally trivial change in the network, in line with my conjecture about transparency (though this is not much of a test of that conjecture).

In paradox 1, with the classically consistent initial state $|\psi\rangle=|0\rangle$, the kinematical consistency condition (12) is satisfied if the density operator $\hat{\rho}$ of the chronology violating bit is

$$\hat{\rho}=p|0\rangle\langle 0|+(1-p)|1\rangle\langle 1| \tag{32}$$

for any real p with $0 \le p \le 1$. The maximum entropy rule therefore sets $p=\frac{1}{2}$, so both $\hat{\rho}$ and the density operator of the output are $\frac{1}{2}\hat{1}$. It is encouraging that this state is continuous with the output state (14) for generic input.

In paradox 2, the maximum entropy rule sets $\lambda=0$ in (19), which again gives a maximally mixed state $\frac{1}{2}\hat{1}$ for the chronology-violating bit, though in this case the output state does not depend on the supplementary data. In paradox 3 no supplementary data are required. Now I can return to paradox 4.

If f is invertible the maximum entropy rule places the n chronology-violating bits in the maximally mixed state $2^{-n}\hat{1}$, and the output state has the same form. Once again we see a striking qualitative difference between the behavior of a quantum computational network and that of its classical analogue. In the classical case the output of the network of Fig. 5 is certainly x'. In the quantum case, provided that f is an invertible function, the output is merely a random number chosen from $\mathbb{Z}_{2^n}$. What is significant here is not the randomness *per se* but that the output is simple [20] and is unrelated to the value x' that would conflict with the evolutionary principle.

But things may be different if f is not invertible. [Note that the gate G_f is reversible, and its evolution (8) unitary, whether or not f is invertible.] If there are idempotent elements of degree 2 or more, then there are states in which the outcome x' has zero probability, though that is no guarantee that the idempotent elements themselves are not knowledge laden. But the paradox manifests itself in its severest form if the only repeated element in the sequence $\{f(x)|x\in\mathbb{Z}_{2^n}\}$ is x', the unique fixed point, and there is no other idempotent element. In that case for some x_0 the sequence $\{f^{(i)}(x_0)\}$ is just $\mathbb{Z}_{2^n}$ in a different order. There can be no $y\in\mathbb{Z}_{2^n}$ such that $f(y)=x_0$, so (27) implies that $p_{x_0}=0$, which in turn implies that $p_x=0$ for every x other than the fixed point x'. The solution is therefore unique so no supplementary data are required. The only state that satisfies the kinematical consistency condition (24) is the one we had hoped would be avoidable, namely $|x'\rangle\langle x'|$.

One way out would be for the evolutionary principle to place constraints on the initial data. Of course it does that anyway, but these would be additional "retrospective" constraints of the same type that classical physics imposes in paradoxes 1–3. I have already said that such constraints would not self-evidently be unphysical, but in any case they may be unnecessary. To see why, we must consider the internal degrees of freedom of the chronology-violating region.

If these degrees of freedom were strictly isolated from the outside, the maximum entropy rule would place them in the maximally mixed state. In practice they would be to some extent coupled to any time-traveling bits in the region, and would be a source of noise in the computation. Suppose that this noise is such that over the period of the journey from the future of gate G_f to its past there is a small probability ε of a transition away from the correct computational basis state (which in the absence of noise would be stationary during that period), and that all erroneous states are equally likely (this is an unrealistic simplification, but correcting it would not significantly affect the result). Then the true kinematical consistency condition would not be (24) but

$$(1-\varepsilon)\mathrm{S}_f\hat{\rho}+\varepsilon 2^{-n}\hat{1}=\hat{\rho}\ . \tag{33}$$

Solutions of (33) still have to have the diagonal form (26), but now, instead of (27),

$$p_x=\varepsilon 2^{-n}+(1-\varepsilon)\sum_{\{y|x=f(y)\}}p_y\ . \tag{34}$$

The solution of (34), if f has the form that we are considering so that

$$x'=f^{(2^n-1)}(x_0)\ , \tag{35}$$

is

$$\begin{aligned} p_{f^{(i)}(x_0)}&=2^{-n}[1-(1-\varepsilon)^{i+1}]\quad (0\le i\le 2^n-1)\ ,\\ p_{x'}&=2^{-n}\left[\frac{1-(1-\varepsilon)^{2^n}}{\varepsilon}\right]\ . \end{aligned} \tag{36}$$

Thus for large n and fixed ε,

$$p_{x'}\sim\frac{2^{-n}}{\varepsilon} \tag{37}$$

and the intended output x' of the network is washed out by noise. Because of the amplification of noise in this computation (a phenomenon which does not occur in any of the other computations we have been considering), a network intended to find an n-bit fixed point would tend to fail when $\varepsilon\gtrsim 2^{-n}$, though with repeated trials and error correction it would remain faster by a large fixed factor ε^{-1} than the classical method of repeatedly applying f until the fixed point is reached.

I conjecture that generically the evolutionary principle can be satisfied without constraining the initial conditions, by setting the supplementary data for the internal degrees of freedom of chronology-violating regions.

DISCUSSION

What happened to the contradictions?

Why are the intuitive arguments wrong? What has happened to the contradictions which, according to both intuitive reasoning and classical analysis, rule out certain initial data in paradoxes 1–3?

Each contradiction arose as a consistency condition for information traveling on a loop in spacetime. This amounted to the requirement that a certain dynamical evolution operator have a fixed point, that fixed point being the state of the information on the loop. We have seen that in quantum mechanics there is always such a fixed point. It is also easy to see that the evolution of a classical system generally has no such fixed point. The negation operation, as in paradox 2, is an obvious example. Admittedly, discrete observables in classical physics, such as those of a classical computer, are always approximations, but as an example of a continuous classical system consider the position of a particle in uniform motion on a circle. The evolution of that system over any time interval that is not a whole number of periods of the motion does not have a fixed point. The quantum version of that same evolution has many fixed points, for instance, the eigenstates of angular momentum.

Let us examine the history of what we now know would really happen in paradoxes 1–3. In paradox 1 a bit approaches the chronology-violating region with the value 1. It encounters an older version of itself which according to (13) is in the maximally mixed state $\frac{1}{2}\hat{1}$ in which the values 0 and 1 occur with equal weight. The younger bit measures the older one, replacing its own value 1 with the negation of the measured value. It thereby enters the very same state $\frac{1}{2}\hat{1}$, since an equally weighted mixture of $|0\rangle$ and $|1\rangle$ is the same as an equally weighted mixture of $|1\rangle$ and $|0\rangle$. After it has gone back in time it plays the role of the older version and finally (14) it emerges, still in the state $\frac{1}{2}\hat{1}$. In paradox 2 there is a similar history, except that the younger bit replaces its own value with the negation of the measured one regardless of what its own value was.

The history of paradox 3 is no longer quite equivalent to paradox 1. The particle [21] is initially on a trajectory that would take it to an earlier time if nothing intervened. It then encounters an older version of itself, in a mixed state of being present and absent. Consequently the younger version enters a mixed state of being prevented and not prevented from going back in time, which is exactly what is needed later along its trajectory when, as the older version, it is in a mixed state of being present at the earlier time and not being present. Finally in the unambiguous future there is a mixed state (23) of there being two particles present and none, with no admixture of 1-particle states, because the younger version of the particle had succeeded in going back in time if and only if the older one had failed.

Conflicting predictions of rival "interpretations" of quantum theory

The above descriptions of the histories of paradoxes 1–3 are in terms of states. They do not say what happens to the actual particles, bits and observers that participate. To do that we need to interpret the states as descriptions of "what happens." Chronology violations would provide a new class of situations in which there would be a large experimentally detectable difference between different "interpretations" [22] of quantum theory. To understand why, let us look more carefully at what it means under various versions of quantum theory for a system to be "in a mixed state."

In unmodified quantum theory (i.e., under the Everett interpretation) individual systems have multivalued observables. For example, if a system is in a generic mixed state $\hat{\rho}$, observables that commute with $\hat{\rho}$ have one realized value for each nonzero eigenvalue of $\hat{\rho}$. It is often convenient to refer to the different values as being held in different coexisting Universes. This terminology is usually the best way of expressing in ordinary language how the multiple coexisting values of separate observables may be correlated. However, it is not capable of expressing everything that can happen to quantum systems; only the quantum formalism can do that.

Under the statistical "interpretation" the density operator is deemed to describe not an individual system but a fictitious statistical ensemble of "identically prepared" systems. The ensemble approximates the behavior of actual systems in that an actual system is taken to be a typical element of the ensemble.

In "collapse" (including "dynamical collapse") theories the density operator of an individual system immediately before an observation summarizes the probabilities of the outcomes of all possible observations on that system. But immediately after the observation the state of the system is no longer that density operator but has changed to an eigenstate of the observable whose value has been observed.

In "pilot wave" hidden variable theories the density operator again refers to an individual system. It contains local information about the nonlocal pilot wave. At any instant each observable of the system has a single value, as does the outcome of any measurement of the observable. All the values are fully determined by the values of hidden (unobservable) variables.

Now recall the consistency condition for the evolution round a closed timelike line. In the quantum case I have taken it to be that the density operator of each chronology-violating bit must return to its original value at a given event, as expressed by (15). That is the correct condition under the unmodified quantum formalism, *but it is either wrong or insufficient under every other version of quantum theory,* just as under classical physics.

That is because in every other version of quantum theory observables are at most single valued [25]. This single valuedness is a kinematical property which imposes its own condition for consistency, namely that all observables that possess values at a given event on a closed timelike line must return to those values if the quantum system travels round the line. In the Everett interpretation it is only the state, which describes, roughly speaking, a collection of values taken as a whole, which must be unchanged after passage round a closed timelike line.

For example, consider paradox 3 under the statistical "interpretation." When the younger and older versions of the observer encounter each other, the older one is in a mixed state of being present and absent. This translates unproblematically under the statistical "interpretation" into a statement about an ensemble: in half of the elements of the ensemble there is such an encounter and in the other half there is not. But what happens next? In those elements of the ensemble in which there was no encounter, the younger version of the observer travels back in time, and experiences the encounter at the same event *in the same element of the ensemble* in which the encounter did not happen—a contradiction, notwithstanding that the density operator has returned to its original value. Under the statistical "interpretation" each element of the ensemble must individually satisfy a consistency condition which, in this case, is simply the classical one.

Under "collapse" theories the observer and all sufficiently frequently observed observables take single-valued routes just as in classical physics (not necessarily the same routes, but that does not help) and must therefore also obey classical consistency conditions. In pilot wave theories there is a similar condition for each observable in addition to the condition, corresponding to that in ordinary quantum theory, obeyed by the pilot wave.

Under the Everett interpretation the history of paradox 3 is quite faithfully described in "multiple universe" terminology: In all universes the observer approaches the chronology-violating region on a trajectory which would go back in time. But only in half of them does the observer remain on that trajectory, because in half the universes there is an encounter with an older version of the observer after which the younger version changes course and does not go back in time. After that, both versions live on into the unambiguous future. In the other half there is no encounter and the observer does go back in time and changes the past (i.e., causes it to be otherwise than is accurately recorded in that observer's memory). Like many quantum effects this is counterintuitive at first, but it is consistent and nonparadoxical. The key thing to bear in mind when trying to visualize it is that in half of the universes (let us call them the "*A* universes") the encounter happens and in the other half (the "*B* universes") it does not happen. This is illustrated in Fig. 6. In the *A* universes an observer appears "from nowhere" (no one having embarked on a chronology-violating trajectory in that universe) and in the *B* universes an observer enters the region and disappears "into-nowhere" (since no one has emerged on the chronology-violating trajectory in that universe). But of course it is not really "from nowhere" and "into nowhere," but from and into the other universes. The final state in each *A* universe has two versions of the observer, with different ages, the older one having started life in a *B* universe. In the final state of each *B* universe the observer is absent, having traveled into an *A* universe.

Here is a clear difference between what unmodified (Everett) quantum theory and others predict about the physical effects of chronology violation. The unmodified theory predicts that under the conditions of paradox 3 there will (from the point of view of an external observer) be a probability $\frac{1}{2}$ that the final state will contain two versions of the time-traveling entity (observer, particle, etc.) and a probability $\frac{1}{2}$ that it will contain none. Every other version of quantum theory predicts that it will certainly contain exactly one—even if this requires the intervention of apparent statistical flukes (presumably more likely than "interference with free will") to prevent certain initial data from being prepared, and there would be no such effect if the apparatus were prepared in one of the classically permitted initial states.

Closed timelike lines would provide "gateways" between Everett universes. This highlights one respect in which the "multiple universes" terminology is inadequate. The time-traveling observer never encounters any barrier or locally distinguished boundary between one universe and another. It is only ever an approximation to speak of things happening "in a universe." In reality the "universes" form a part of a larger object which has yet to be given a proper geometrical description but which, according to quantum theory, is the real arena in which things happen. Figure 6 shows that in the presence of chronology violations all the Everett universes are connected into a single manifold; however, that manifold and its geometry do not form a spacetime in the usual sense because the contents of each branch are constrained to be identical in the past of a certain horizon, as shown. It cannot be made into a spacetime by identifying the two branches to the past of the horizon because that makes the manifold pathological (non-Hausdorff) at the horizon (cf. the discussion in Ref. [15]).

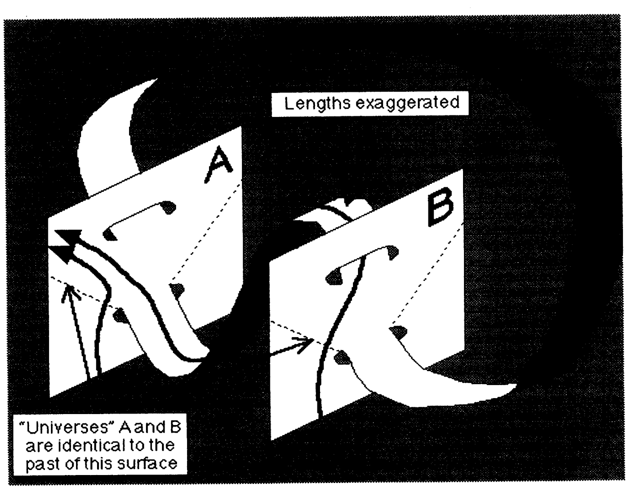

FIG. 6. Spacetime as perceived by observers in paradox 3.

Pure states evolve into mixed states

According to quantum mechanics, in regular chronology-respecting spacetimes a system initially in a pure state always remains in a pure state; moreover, the state of a sufficiently large system, if necessary the contents of a maximal spacelike hypersurface, is always pure.

Hawking [26] has shown that this is no longer true in spacetimes containing evaporating black holes. There, pure states evolve into mixed states. That is because the degrees of freedom on a maximal spacelike hypersurface to the future of the evaporated black hole remain in an entangled state with the degrees of freedom that were "lost" in the black hole. The joint state of the degrees of freedom on the hypersurface and those in the black hole remains pure.

In chronology-violating spacetimes there is a similar effect. We have seen that a pure state in the unambiguous past may well evolve into a mixed state in the unambiguous future. The degrees of freedom in the unambiguous future remain in an entangled state with earlier versions of themselves, i.e., with the supplementary data in the chronology-violating region. However, unlike in the black-hole case, even the combined future-plus-supplementary state can be mixed.

We must therefore reconsider the idea that the Universe as a whole may be described by a density operator rather than a state vector. I have elsewhere [23] been rash enough to describe this idea as a "heresy" and indeed under chronology-respecting circumstances there is a decisive objection to it: the additional information in a "density operator for the Universe," compared with any of its eigenstates considered as a pure state, is strictly unobservable as I shall now prove.

A spacetime satisfies the *stable causality condition* if it is chronology-respecting under every infinitesimal variatin of its metric. Since there would be no experimental way of detecting that a chronology-respecting spacetime violated this technically stronger condition, we may restrict attention to spacetimes that satisfy it. In all such spacetimes there is a function whose gradient is everywhere timelike [3]. This function can serve as a global time coordinate t. Suppose that the contents of a chronology-respecting spactime are described by a density operator $\hat{\rho}(t)$. The evolution of $\hat{\rho}(t)$ will be unitary

$$\hat{\rho}(t)=U(t)\hat{\rho}(0)U^{\dagger}(t) \quad . \tag{38}$$

This implies that $\hat{\rho}(t)$ has the same spectrum, say $\{p_i\}$, as $\hat{\rho}(0)$, and has eigenstates $\{|\psi_i(t)\rangle\}$ that are related to those of $\hat{\rho}(0)$ by

$$|\psi_i(t)\rangle=U(t)|\psi_i(0)\rangle \quad . \tag{39}$$

$\hat{\rho}(t)$ therefore describes a collection of "universes" (each one itself consisting of multiple universes under the Everett interpretation), one for each nonzero p_i. Each evolves precisely as if the others were absent and it had a pure state $|\psi_i(t)\rangle$. This is quite unlike the Everett multivaluedness caused by the linear superposition of components of a state vector, which is detectable through interference phenomena. Thus the cosmology described by $\hat{\rho}(t)$ contains a multiplicity of mutually disconnected and un-observable entities and is vulnerable to the "Occam's razor" argument that is sometimes erroneously leveled against the Everett interpretation.

But in the presence of closed timelike lines the evolution with respect to an external time coordinate is no longer necessarily unitary as in (38). Nor is it even (as it remains in the black-hole case) necessarily the restriction to a subsystem of a unitary evolution in a larger system; for all such evolutions are described by linear superscattering operators whereas the evolution to (14), for example, is nonlinear in the initial density operator $|\psi\rangle\langle\psi|$. Therefore in principle it might be possible to detect experimentally the difference between distinct density operators with identical eigenstates, so the "Occam's razor" argument no longer necessarily holds.

Note that the hypothesis that the Universe was in a pure state at its past boundary remains viable and (for at least the reason which I am about to discuss) attractive.

Time-reversal invariance and the second law of thermodynamics

An evaporating black hole is itself time asymmetric. It starts as regular matter and ends in a naked singularity. But a chronology-violating region need not be. For example, the spacetime of Fig. 1(b) is time symmetric. Moreover all the interactions that we have considered are locally unitary and therefore time reversible. Are the phenomena that we have been discussing time reversible also? Suppose that preparing the standard network of Fig. 3 in a pure input state $|\psi\rangle$ gives an output in a mixed state $\hat{\rho}$ when the unitary evolution operator of the gate G is U. One might expect that when the evolution operator of G is $U^{\dagger}$ and the input is prepared in the state $\hat{\rho}$ the output would be in the pure state $|\psi\rangle$.

But one would be mistaken. Mixed states never evolve into pure states, even in the presence of chronology violations. Although all the interactions are reversible, there is an implicit requirement on correlations in the initial state which does not apply to the final state: when we speak of preparing the first m bits in a state $\hat{\rho}_1$ we implicitly mean that the joint density operator is $\hat{\rho}_1 \otimes \hat{\rho}_2$ for some state $\hat{\rho}_2$ of the last n bits. We do not consider it possible that $\hat{\rho}_1$ could be prepared as a restriction $\mathrm{Tr}_2\hat{\rho}$ of some general entangled density operator $\hat{\rho}$ of the $m+n$ bits entering the gate, because the last n of those bits do not yet exist at the time when the first m must be prepared. But when we speak of the *output* having a certain density operator $\hat{\rho}_{\mathrm{out}}$ we mean precisely that $\hat{\rho}_{\mathrm{out}}=\mathrm{Tr}_2\hat{\rho}$ where $\hat{\rho}$ is an entangled joint state of all $m+n$ bits when they emerge from the gate.

This assumption about the form of the initial and final states of interacting systems is a general one that underlies quantum measurement theory, namely that physically separated systems get into entangled quantum states only *after* interacting with each other. This is a time-asymmetric assumption. It is related to the second law of thermodynamics, for if a mixed state evolved into a pure state the entropy of the Universe would go down. The usual hypothesis for the initial value of the entropy is that it is zero, which would imply that the initial state of the Universe is pure.

Under quantum mechanics the kinematical consistency condition does not allow any process to "evade" the second law by dumping entropy into the chronology-violating degrees of freedom just before they cease to exist. The entropy functional on the space of density matrices

$$S[\hat{\rho}]=-\mathrm{Tr}(\hat{\rho}\ln\hat{\rho}) \tag{40}$$

has the property of subadditivity

$$S[\hat{\rho}]\leq S[\hat{\rho}_1]+S[\hat{\rho}_2] \quad , \tag{41}$$

where $\hat{\rho}$ is the density operator of a system composed of two subsystems with density operators $\hat{\rho}_1$ and $\hat{\rho}_2$. Consider the effect of the general network of Fig. 3 on the total entropy. The total entropy of the input in the unambiguous past is $S[\hat{\rho}_1]$. I shall not address the question whether it is physically meaningful to evaluate the entropy within the chronology-violating region, but formally, just formally, just before the interaction in the gate G it has become

$$S[\hat{\rho}_1\otimes\hat{\rho}_2]=S[\hat{\rho}_1]+S[\hat{\rho}_2] \tag{42}$$

by virtue of the appearance of the chronology-violating bits. By subadditivity the entropy of the output state satisfies the inequality

$$S[\hat{\rho}_{\mathrm{out}}]\geq S[U(\hat{\rho}_1\otimes\hat{\rho}_2)U^{\dagger}]-S[\hat{\rho}_2] \quad , \tag{43}$$

the last term being a consequence of the kinematical consistency condition, and from (40), (42), and (43),

$$S[\hat{\rho}_{\mathrm{out}}]\geq S[\hat{\rho}_1] \quad . \tag{44}$$

In summary, under quantum theory the second law of thermodynamics holds in the presence of chronology violations, except possibly within the chronology-violating region where its meaning is unclear.

That is not so in the classical case, where the second law can be evaded. Consider, for example, the network of Fig. 1(a) where G is again a measurement gate but this time with the roles of its two inputs reversed relative to the gate in paradox 1; i.e., it effects the evolution

$$|x\rangle|y\rangle \Longrightarrow |x\rangle|x \dot{+} y\rangle \quad (\forall x,y \in \mathbb{Z}_2) \ . \tag{45}$$

The kinematical consistency condition forces the second input bit of the gate to have the same value as the first, so the output is always zero. This evolution evades the second law in that it sets an unknown input value to zero without generating any waste information [27]. The waste information, in the form of the value of the time-traveling bit, has been dumped in degrees of freedom that have ceased to exist by the time of the unambiguous future. Such a network could then be used to construct a "Maxwell demon" device (perpetual motion machine of the second kind) unless somehow the chronology-violating links had an entropy of their own and radiated it away, like black holes do, before they disappeared. There is no apparent mechanism for that—but it is pointless to pursue the matter further since this whole problem arises only under classical physics.

In the quantum case the same network has the same effect *on pure computational basis states* of the input; i.e., it effects the nonunitary evolution

$$|x\rangle \Longrightarrow |0\rangle \quad (\forall x \in \mathbb{Z}_2) \ . \tag{46}$$

This is a novel and interesting phenomenon, but one thing it does not do, since all pure states have zero entropy, is violate the second law. The evolution of all input states satisfies (44) so entropy never decreases.

Global conservation laws

Without quantum effects global conservation laws go wrong at the transitions between chronology-respecting and chronology-violating regions. A particle whose trajectory is a closed timelike line comes into existence at a certain (externally defined) instant and then ceases to exist at a later (externally defined) instant. If a particle travels to an earlier time at which it had already existed, then for a certain period there will be two versions of the particle present whereas in the unambiguous past and future there is only one. If the particle is a baryon, for instance, this means that the total baryon number of the Universe is first increased and then decreased by one. The same could happen to other conserved quantities, such as charge or energy, that the particle can carry.

Friedman *et al.* [5] have shown that all global quantities obtained by integrating a conserved local flux over a spacelike hypersurface (including baryon number and charge, and energy if the region is stationary) remain conserved within a chronology-violating region. This result holds in both classical and quantum physics. It does not necessarily apply to the transition between chronology-violating and chronology-respecting regions, because the constant-external-time hypersurfaces may, as in Fig. 1(b), have instantaneous singularities at such transitions, similar to the singularity at the instant of evaporation of a black hole.

However, if the singularities are sufficiently benign for a computational-network model to be faithful—i.e., on the reasonable assumption that nothing comes out of them that could not in principle be computed in a nonsingular region—then the result of Friedman *et al.* can be extended. In the classical case, no matter what may happen inside the chronology-violating region, provided that the interactions locally respect the conservation laws, all additive conserved quantities will, by the time of the unambiguous future, be restored to the values that they had in the unambiguous past. This may be seen by inspection from Fig. 3: The chronology-violating bits must all return to their original states, and therefore to the original values of all their conserved quantities. And since these quantities are additive, and their total is unchanged by the action of the gate G, that total must be the same in the output state as in the input.

Under quantum mechanics this need no longer be so. We have seen that in paradox 3 a conserved quantity, the particle number, could change in a chronology-violating region and remain unchanged in the unambiguous future, though its expectation value was still conserved. I shall now show that this is true in general. *The expectation value of an additive conserved quantity is globally conserved* between any two spacelike hypersurfaces on which chronology is respected, even if there is chronology violation in the region between the hypersurfaces.

Consider the network of Fig. 3 again. Let

$$\hat{X} = \hat{X}_1 \otimes \hat{1} + \hat{1} \otimes \hat{X}_2 \tag{47}$$

be an additive conserved quantity where $\hat{X}_1$ and $\hat{X}_2$ are its restrictions to the first m and last n bits, respectively. The conservation law is expressed by

$$U^\dagger \hat{X} U = \hat{X} \ . \tag{48}$$

Since the evolution in general maps pure states of the input to mixed states of the output, it is too much to expect all eigenstates of $\hat{X}_1$ to remain unchanged. However, the expectation value of $\hat{X}_1$ does remain unchanged, for in the input state it is

$$\mathrm{Tr}(\hat{\rho}_1 \hat{X}_1) \tag{49}$$

and in the output state it is

$$\begin{aligned} \mathrm{Tr}_1[\mathrm{Tr}_2(U(\hat{\rho}_1\otimes\hat{\rho}_2)U^\dagger)\hat{X}_1] &= \mathrm{Tr}(U(\hat{\rho}_1\otimes\hat{\rho}_2)U^\dagger \hat{X}) \\ &\quad - \mathrm{Tr}_2[\mathrm{Tr}_1(U(\hat{\rho}_1\otimes\hat{\rho}_2)U^\dagger)\hat{X}_2] \ . \end{aligned} \tag{50}$$

The conservation law (48) and the cyclic invariance of the trace simplify the first term on the right-hand side of (50) and the kinematical consistency condition (15) simplifies the second term. Thus the expectation value of $\hat{X}_1$ in the output state is

$$\mathrm{Tr}[(\hat{\rho}_1\otimes\hat{\rho}_2)(\hat{X}_1\otimes\hat{1}+\hat{1}\otimes\hat{X}_2)] - \mathrm{Tr}(\hat{\rho}_1\hat{X}_2) = \mathrm{Tr}(\hat{\rho}_1\hat{X}_1) \tag{51}$$

which is unchanged from (49).

Stability of the effects

The effects of quantum multivaluedness usually appear experimentally in the form of interference phenomena.

The detection of interference in a quantum system requires the maintenance of quantum coherence in that system, which often becomes prohibitively difficult as the system in question becomes large and complicated. That is perhaps why quantum theory is often mistakenly described as a theory only of "microscopic" systems. But the effects which I am describing, though they are fully quantum mechanical in that they have no classical analogues, are not interference phenomena and do not require the maintenance of coherence. They are stable at the macroscopic level.

Unlike interference effects, the step-by-step histories of paradoxes 1–4 are quite insensitive to being continuously watched. For example, suppose that in paradox 3 an external observer makes a measurement of whether or not an older version of the particle (or time traveler) has on this occasion appeared, before the moment when the two versions would begin to interact. Then, in terms of the simulation, the density operator of the older bit loses its off-diagonal elements and coherence is lost. But those off-diagonal elements were zero anyway. No interference can now be detected between the two branches of the history corresponding to the presence and absence of the older version. Detecting such interference would require the measurement of observables incompatible with the "occupation number" observable that has just been measured. But the effects that we are interested in do not involve the measurement of such an observable. For example, in order to perform the test of the Everett interpretation, we need only observe whether, after the experiment is over, the younger version is present or absent. That measurement is compatible with the earlier one, though their outcomes are correlated and those correlations cannot be calculated from the separate density operators of the two versions (which are both $\frac{1}{2}\hat{1}$). To calculate the outcomes, note that the performance of these two measurements is equivalent to measuring, after the interaction between the versions has or has not taken place, a nondegenerate observable with eigenstates

$$|00\rangle,\ |01\rangle,\ |10\rangle,\ \text{and } |11\rangle \tag{52}$$

[cf. (21)]. The probability of having observed the presence of exactly one version is

$$\tfrac{1}{4}\mathrm{Tr}[(|01\rangle\langle 01|+|10\rangle\langle 10|)\times(|00\rangle\langle 00|+|11\rangle\langle 11|)]=0\ . \tag{53}$$

At least, that is what is predicted under the Everett interpretation. All other versions of quantum theory agree with classical physics that the number of versions observed will be 1 with certainty. This follows from what I have said, but perhaps it is worth elaborating the prediction in the case of "collapse" theories. Under such theories the first observation will cause the state of the older version to collapse to one of the two states $|0\rangle$ or $|1\rangle$ representing the absence or presence respectively of a time-traveling particle. Call that state $|x\rangle$. Then the interaction (5) in the gate will cause the younger version to emerge in the state $|x \dot{+} 1\rangle$. That is an eigenstate of the observable measured in the second observation, so the outcome of the second observation will certainly be $x \dot{+} 1$ and the second collapse will have no effect. Since the state does not evolve other than inside the gate and during measurements, the kinematical consistency condition is no longer (22) but simply

$$|x\rangle = |x \dot{+} 1\rangle\ , \tag{54}$$

a contradiction. Therefore, under "collapse" theories the interaction specified in paradox 3 can never happen, and the original particle will with certainty survive unaccompanied.

On the subject of allowed particle numbers in the unambiguous future of paradox 3, the predictions of all versions of quantum theory are stable against stray perturbations or measurementlike interactions with the environment. And those predictions all agree, except those of unmodified (Everett) quantum theory which are qualitatively different. All we need in order to perform a crucial experimental test that would settle the interpretation controversy is access to a closed timelike line.

Violation of the correspondence principle

For quantum and classical physics to make qualitatively different predictions about the behavior of a macroscopic object, especially if those predictions are stable, is a violation of the correspondence principle. It follows that if closed timelike lines can exist on a large enough scale to accommodate macroscopic time-traveling objects, the correspondence principle is false.

Relativity of probability

According to (23) an external observer viewing repeated trials of paradox 3 would, on about half of the occasions, see the participating observer disappear forever, and on the other half, see an extra instance of that observer appear. But the experience of the participating observers themselves would be quite different. From their point of view (see Fig. 6) it is *certain* that the final state contains two versions of themselves, and *impossible* that it contains none.

This is a matter of relative probabilities. Equation (23) says that the participating observer is not present in half of the universes in the future. Therefore, although the absolute probability of there being two versions of the participating observer in the final state is

$$\tfrac{1}{2}\mathrm{Tr}[|11\rangle\langle 11|(|00\rangle\langle 00|+|11\rangle\langle 11|)]=\tfrac{1}{2}\ , \tag{55}$$

the relative probability of that outcome, given that the participating observer is present to observe it, is

$$\frac{\tfrac{1}{2}\mathrm{Tr}[|11\rangle\langle 11|(|00\rangle\langle 00|+|11\rangle\langle 11|)]}{\tfrac{1}{2}\mathrm{Tr}[(\hat{1}\otimes\hat{1}-|00\rangle\langle 00|)(|00\rangle\langle 00|+|11\rangle\langle 11|)]}=1\ . \tag{56}$$

The only other place in physics where one calculates probabilities relative to the very existence of the observer is the anthropic principle [28]. The weak anthropic principle is the statement that for the purposes of testing probabilistic theories it is wrong to calculate the absolute probabilities that measurements will have given outcomes, but correct to calculate relative probabilities given the existence of an observer. But it is very difficult to make that statement precise, and its status is still controversial. For example, would a rational observer be right to choose a course of action that had a low probability of yielding a very high reward but would otherwise cause instant painless death (assuming that the observer places sufficiently little value on the effect of this on others)? Or would such an observer be right to argue "admittedly if I did this I should experience only those universes in which I should be rewarded, but I should still know of others in which I had ceased to exist, and I consider that cessation of existence as something that would have happened 'to me,' and as such it is undesirable"?

The present example is more clear-cut. In classical physics the only way in which an observer can permanently cease to exist is to die. That is, there is always a particular event at which the information that constituted that observer is destroyed or inactivated, and as I have said the observer could contemplate that event prospectively and might reasonably place a value on it. In paradox 3 there is no such event. Although the observer eventually no longer exists in half the universes, there is no uniquely defined last event at which that observer existed in those universes (though there is a limiting event at which the observer can no longer be seen from the unambiguous future in those universes), and the observer's own experience is continuous. Neither subjectively nor objectively is there any destruction or inactivation of the observer, merely motion from one place to another. Yet the observer's subjective probabilities may be utterly different from the objective ones.

An entirely new type of experience made possible by this effect might be called "asymmetric separation": Adam and Eve live on a desert island with a time machine. Eve wants more company so she decides to use the time machine to create another version of herself. From her point of view her plan carries zero risk. She will with certainty end up on the same island with Adam and another version of herself. The only thing that she cannot predict is whether she will actually have to travel in the time machine, or whether, as she walks towards it, an older version of her will step out. But from Adam's point of view Eve is gambling. There is only a 50% chance that the gamble will come off favorably as she expects, and there is also a 50% chance that she will be separated from him and never return. Thus it is possible for Eve to be separated from Adam without Adam being separated from Eve.

After Eve disappears, Adam can follow her and certainly find her, just as if they had gone through together, provided that he does so before the chronology-violating region ceases to exist. Both of them will then experience only a world containing two Adams and two Eves. It would also be possible for Adam and Eve to go through the process independently, in which case there would be four equally likely outcomes from the point of view of an external observer, with Adam and Eve independently absent, or present in two versions. From Adam's or Eve's point of view, each of them would be certain of existing in two versions, and be equally likely to be or not be accompanied by two copies of the other.

What is the difference between going through "independently" and going through "together"? It has nothing to do with whether they follow similar paths (they may even take topologically inequivalent routes), nor what period of time elapses between their going nor (as in science fiction stories) whether or not they hold hands. What matters is the interaction that occurs between the *older and younger versions* of who- or whatever goes through. In order to create reliably a copy of herself, Eve must arrange for that interaction reliably to be that she will travel in the time machine if no older version of her comes out, and not travel if an older version of her does come out. If Adam traveled if and only if no version of *him* came out he would be going through independently. To be sure of creating coexisting copies of both of them, they must ensure that they will both travel on the time machine if and only if neither comes out (it will then not happen that exactly one of them comes out). A third possibility is that while Eve creates a copy of herself in the way just described, Adam travels on the time machine if and only if *Eve does* come out. Each of them will then be certain of being separated from the other, having exchanged the other for another version of themselves. An external observer will then perceive it to be equally likely that there will be two Adams or two Eves, and impossible that the island will remain unoccupied.

One can speculate what might happen if we one day discover, or create, a closed timelike world tube large and regular enough to accommodate safely the solar system, and we have the means to move the solar system into it. Then we should be in a position to create any number of copies of the solar system. Repeating the process n times would create 2^n solar systems, and the process would succeed with subjective certainty for every member of the human race (assuming that they were all on board). The absolute probability that the solar system would still exist afterwards would fall to 2^{-n}, but presumably that would not matter because the probability that any of the inhabitants would be destroyed would be zero.

Perhaps this is the explanation of the fact, which is sometimes considered puzzling ([29] and Chap. 9 of [28]) that we have observed no intelligent life other than *homo sapiens.* If the effect is typically used by civilizations who are able to, and the time taken to reach the required technological level is small compared with the time taken for a civilization to spread across the galaxy, then it would be very unlikely that any young civilization such as ourselves would yet have observed other intelligent life even if civilizations come into existence quite frequently in the galaxy.

Two civilizations with conflicting plans for the resources of the galaxy could, by cooperating, use the effect to get out of each other's way. Each could have the galaxy to itself without ceding any resources other than those already occupied by the other.

It would also be possible in principle for the very early Universe as a whole to have undergone such a process spontaneously. Depending on what the interactions were, this could cause all sorts of symmetry breaking.

Implications for the quantum formalism

Conventionally in the "quantization" of classical systems the domain, or base space (such as spacetime), of the classical observables is unchanged. Only the range, or configuration space, changes as the classical real-valued observables are "promoted" to Hermitian operators. In semiclassical approximations to quantum gravity both the configuration space and the base space change; for example, the topology of the spacetime in which a black hole evaporates cannot be taken to be the same as that of the classical black-hole spacetime. In general, the manifold in which the fields propagate in semiclassical quantum gravity depends on the quantum state and on the interactions to which the fields are subject.

The same effect occurs, for a different reason, when chronology is violated. In the presence of the interactions of paradox 3 in the quantum case the spacetime of Fig. 6 replaces the classical spacetime of Fig. 1(b) as the set of events through which the world lines of the particles may move. The same would presumably be true (because of the stability of the effect) if the classical carriers of this model were replaced by realistic particle states of a quantum field, carrying the bits as internal degrees of freedom. It must be borne in mind when picturing the physics of time travel through diagrams such as Fig. 6 that the spacetime would, in general, be different if the interaction between the bits were different.

Both in the evaporating black-hole case and in the cases discussed in this paper the background spacetime has been identified *ad hoc.* The construction of a general mathematical framework for this is an important task for the future and is, among other things, likely to be a prerequisite for a full quantum theory of gravity.

Chronology violation falsifies some of the assumptions that are built into some formulations of quantum theory. However, the quantum formalism is still a complete, consistent, and viable framework for physics whether or not there are closed timelike lines. Proving this proposition rigorously is another interesting task for the future, but I see no reason to expect the problems involved to be other than technical. Let us take stock of what has to be dropped and what remains in the formalism.

The most important thing to go is unitary evolution. But unitarity is not violated so badly as to contradict the probability calculus. Probabilistic predictions can still be made in the usual way, provided that where necessary one makes them relative to the experience of the relevant observers.

One can still construct a Hilbert space of global quantum states on any given maximal spacelike hypersurface, and these are related to the states of local systems on that hypersurface in the usual way. The change in particle numbers as particles loop back in time requires the space of quantum states to accommodate variable numbers of particles (i.e., to be a Fock space) even for nonrelativistic and noninteracting systems. Insofar as standard methods of constructing such spaces (e.g., quantization) work when chronology is respected, they work when it is violated. The spaces corresponding to different hypersurfaces will not be unitarily equivalent, nor can the dynamics be described in the usual way as motion on a global Hilbert space. Neither the Schrödinger nor the Heisenberg picture can be used to describe quantum dynamics globally, but either of them can still be used locally, as in this paper. Echeverria, Klinkhammer, and Thorne [30] have successfully applied the path-integral formalism to simple chronology-violating quantum systems, and that is also likely to be the correct approach for quantum field theory.

Cloning quantum systems

In quantum theory if chronology is respected it is impossible to *clone* a physical system. That is, given an original quantum system which is not known to be in a state from a known basis, it is impossible to place another system, the clone, in the same state as the original while leaving the original in that state and to know that one has done so. By the "same" state we mean, as usual, the same under a given one-to-one correspondence between the states of the original and the clone. The result holds whether by "the state" of the system we mean a pure state of the system alone, a pure state of the system and all its correlations, or the density operator of the system alone. By "impossible" we mean that no process can have a nonzero probability of achieving the stated effect.

This result is well known (for a recent treatment, see Ref. [31]). Let me rederive it in a way that shows how it breaks down when chronology is violated. To clone a system in an unknown state $\hat{\rho}$ means to effect an evolution of the form

$$\hat{\rho}\otimes\hat{\rho}_0\otimes(\hat{1}-\hat{P})\Longrightarrow p(\hat{\rho})\hat{\rho}\otimes\hat{\rho}\otimes\hat{P} \\ +(\text{other terms})\otimes(\hat{1}-\hat{P}) \tag{57}$$

with $p(\hat{\rho})>0$, where $\hat{\rho}_0$ is some fixed initial density operator for the system that is to become the clone and $\hat{P}$ is a projection operator for the "yes" state of a 2-state system that indicates whether or not the process has succeeded.

Let $\hat{\boldsymbol{\rho}}$ be the density operator of the universe as a whole before the evolution (57). That is,

$$\hat{\rho}=\mathrm{Tr}_{\neq 1}\hat{\boldsymbol{\rho}} \ , \tag{58}$$

where $\mathrm{Tr}_{\neq 1}$ denotes the trace over all but the subspace of the original system. If $\mathbf{U}$ is the unitary evolution operator for the Universe for the period during which (57) happens, we have

$$\mathrm{Tr}_{\neq 1,2}(\mathbf{U}\hat{\boldsymbol{\rho}}\mathbf{U}^{\dagger}\hat{P})=p(\hat{\rho})\hat{\rho}\otimes\hat{\rho} \ , \tag{59}$$

where $\mathrm{Tr}_{\neq 1,2}$ denotes the trace over all but the subspaces of the original system and the clone. In view of (58) the right-hand side of (59) is a nonlinear function of $\hat{\boldsymbol{\rho}}$. But the left-hand side, which represents the most general possible evolution, is linear in $\hat{\boldsymbol{\rho}}$.

On the assumption that $\hat{\boldsymbol{\rho}}$ is not initially, by some conspiracy in the initial conditions, the appropriate nonlinear function of $\hat{\rho}$ (in which case the Universe would in effect already contain the required clone), it will not become one and therefore cannot satisfy (57).

It is this assumption which can be falsified by chronology violation. If chronology is respected, such a "conspiracy in the initial conditions" would violate both the evolutionary principle and the usual assumptions of measurement theory. But if chronology is violated, supplementary data are introduced. Because several instances of the same bit may exist at the same time, the density operator of the Universe on a spacelike hypersurface that passes through the chronology-violating region describes more degrees of freedom than the density operator in the unambiguous past; and those additional degrees of freedom are partially constrained by the kinematical consistency condition to be functions of the existing ones, in this case of $\hat{\rho}$. For example, if the system were inert for a period while it looped n times from "after" to "before" a given hypersurface, then $\hat{\rho}$ on that hypersurface would be constrained to contain $n+1$ factors of $\hat{\rho}$—something that dynamics alone (i.e., without constrained initial or supplementary data) could never achieve. If chronology is violated $\hat{\rho}$ is kinematically constrained to be some nonlinear function of $\hat{\rho}$, and that is why a dynamical evolution of the form (57) is not ruled out.

Implications for measurement theory

The cloning phenomenon can be harnessed to permit new types of quantum measurement. Perhaps the most striking one is that the observables of a system are no longer the only quantities that can be measured. With the help of closed timelike lines it is also possible to measure the state (pure or mixed) of a system. One simply creates one or more clones and performs measurements on those. Measurements of the state with arbitrary precision are thereby made possible; however, there is a trade-off between the precision of the measurement and the probability of not losing the system into another universe—unless one is willing to clone oneself as well.

The unattainability of (57) is sometimes cited as being responsible for the impossibility of using nonlocal quantum correlations for signaling. I leave it as an exercise for the reader to verify that signaling between noninteracting systems remains impossible in the presence of closed timelike lines by proving that (a) if a quantum computational network consists of two subnetworks that do not interact, any change made in either subnetwork is denotationally trivial for the other, even if the inputs of the two subnetworks are in an entangled state, and (b) this remains true relative to the subjective experience of a cloned observer of either of the systems.

When a system is cloned, its nonlocal correlations are not. Only one system (which we may call the "original" one) remains in an entangled state with accessible versions of distant systems. The clones are in entangled states with instances of the distant system in other universes, but not with the original system nor with each other.

Spacetimes without chronology-respecting regions

If a spacetime has no chronology-respecting region, the maximum entropy rule would require its contents to be at an absolute maximum of entropy. They would have to be in the mixed state $(1/N)\hat{1}$ where N is the dimensionality of the state space of the contents of spacetime. Hawking has suggested for other reasons that this might be the state of the Universe even *including* spacetime. However, the objection to mixed states on the grounds of unobservability, which I have argued does not always apply if there are chronology violations, does apply strongly in both these cases.

As I have said, the maximum entropy rule is only one way of implementing the evolutionary principle. We should be extremely reluctant to abandon the evolutionary principle, but there might be other ways of implementing it; however, the evolutionary principle itself makes it very implausible that there is no chronology-respecting region in spacetime: If evolution is to be responsible for the existence of all the knowledge in spacetime, every knowledge-containing region must be contained in the future Cauchy development of a knowledge-free region. But that means that although every observer is on some closed timelike line, no information about the observer travels around the line. Information about an observer would presumably contain nonzero knowledge, so knowledge would then be present everywhere on the closed timelike line, including where it passes through the knowledge-free region—a contradiction. Therefore the closed timelike lines must provide no closed path for knowledge. But once again, although knowledge is not information, a strictly zero flux of knowledge presumably implies a strictly zero flux of information, so there would be no closed path for information either. The knowledge-free region would have to be opaque. But in that case there would be no observable difference between the spacetime in question and one from which every knowledge-free region had been excised. The latter spacetime, if it contained an observer, would contain chronology respecting regions including an unambiguous past boundary with zero-knowledge initial data, just as I have been assuming all along.

This excision is similar to the denotationally trivial transformation which converts a chronology-violating computational network in which there is no closed path for information into a chronology-respecting network. In both cases the formal chronology violation has no physical significance and can be defined away. Perhaps in a future unified theory, chronology will appear as a joint property of all fields and their interactions, not just spacetime.

Implications for the theory of computation

The theory of computation may be conveniently divided into two branches: the *theory of computability* is about what computational tasks can be performed in nature, and *complexity theory* is about what physical resources are required to perform them. Chronology violations would affect both branches. The details are peripheral to the subject of this paper, in which computations are being studied merely as a convenient way of investigating new physical effects. But a few comments are in order.

We have seen many examples of chronology-violating computational networks that are not denotationally equivalent to any chronology-respecting network. *Ipso facto,* chronology violation permits qualitatively new types of computational tasks.

As for complexity theory, it seems likely that many of the effects could be harnessed to allow various computational tasks to be performed with greater efficiency than would be possible without chronology violations. However, it is difficult to give any quantitative statement because it is not yet known what resources would be required to give chronology-violating computational components, such as negative delays, their functionality.

Some of the considerations involved are illustrated by the network shown in Fig. 7(a). N is a general chronology-respecting quantum computational network, not necessarily a gate. This is indicated in the diagram by the representative *spatially* looping path in the right (of which there may be many within N). N could be a general-purpose quantum computer so we may as well suppose that it is one. The line entering N on the left represents any number of input bits. The line leaving N on the left carries at least three items of information: (a) A periodically emitted "halt flag" which has the value 0 while the computation is still running and 1 when it has halted; (2) the output of the computation; (3) some control information for the negative delay, setting it to $-T$, the negative of the time for which N was running.

Like many spatial diagrams of chronology-violating networks, Fig. 7(a) is potentially misleading about the implied physical situation. To avoid being misled one must bear in mind that unlike other components of a computational network, negative delays signify something about how the network is *globally* embedded in spacetime. Although N and the long negative delay may be thought of as being logically consecutive as in Fig.

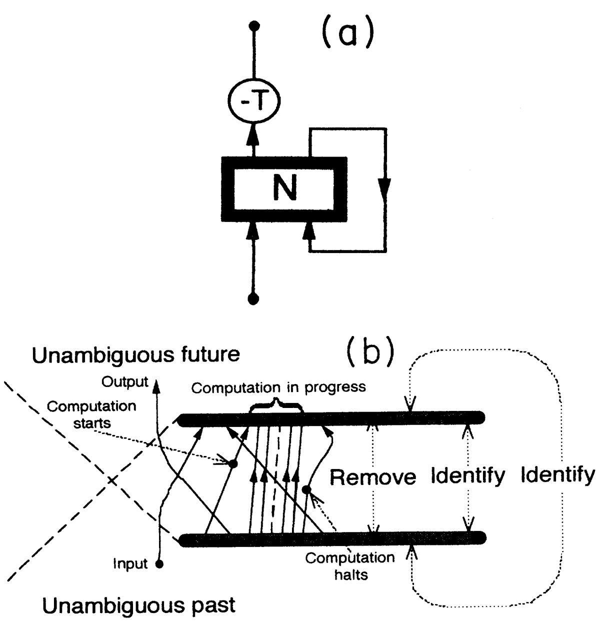

FIG. 7. (a) How to compute anything in no time? (b) Spacetime implementation of Fig. 7(a).

7(a), physically they must be concurrent or interleaved; for since the time-traveling output bits of N must be able to travel to the ambiguous past of the input, the whole computation performed by N must be within or to the unambiguous past of the chronology-violating region of the spacetime. One way of arranging for the effect of this network is illustrated by the spacetime diagram Fig. 7(b). After being given the input, the computer N is put into a spaceship and programmed to fly into the chronology-violating region and remain there until the desired computation has halted, and then to emerge and deliver the output. It starts in the unambiguous past. It is given its input at or near the event where a much older version of it delivers the output. Then, after starting the computation and initially moving rapidly to the right, it slows down and moves only enough to avoid colliding with younger versions of itself. Finally, when the computation halts, it moves rapidly to the left and delivers the output at or near the event where it received the input, and also carries the output on to the unambiguous future.

We see that only a user within the chronology-violating region can actually obtain the output instantly. A user at a distant location must wait for approximately the light travel time to the chronology-violating region and back between preparing the input and receiving the output, but even that user seems to receive the output of an arbitrarily complex computation in a fixed time. Furthermore this method seems to provide an oracle for the halting problem, for the output appears after a fixed time if and only if the computation halts. However, such literally infinite computing power is made impossible by unavoidable physical limitations.

First, if the carrier of the computation (the "spaceship") is to have an information-storage capacity of n bits it must occupy a volume of approximately $n^{3/2}$ Planck units or more because of the bound imposed by black-hole thermodynamics [32]. But the entire computation must take place within the chronology-violating region. Therefore if that region has a finite spatial volume there will be an upper bound on the number of versions of the carrier that can coexist on one spacelike hypersurface in the region, and therefore there will be a limit on the proper time that the carrier can spend inside the region. The limits that this imposes on computations are complicated, since there is a trade-off between the total time available and the memory capacity of the carrier, and an additional complication is that it may be advantageous for some of the instances of the computer to interact with each other; but for our present purpose the important thing is that there are limits.

Second there are the resources required to manufacture the chronology-violating regions by some as yet unknown method.

The Church-Turing principle and the feasibility of time machines

That brings me to the question whether or not closed timelike lines can exist [33] in nature. For all its virtues, the basic method I have used is not well adapted to answering that question. The very universality of the quantum theory of computation means that it does not "know" what raw materials and what interactions are available in nature, because quantum theory itself does not "know" that any more than it "knows" whether quarks or quasars can exist in nature. Such knowledge is the domain of lower level constitutive theories of the Universe. Since we do not yet have the most relevant constitutive theory, namely quantum gravity, we do not know enough to say whether or not closed timelike lines can exist, though they do exist at the level of quantum fluctuations in many current approaches to quantum gravity. In any case the results of this paper show, contrary to what has usually been assumed, that there is no reason in what we currently know of fundamental physics why closed timelike lines should not exist.

If they can exist, can we create them artificially? That is, can time machines be manufactured? On that conditional question we can make more headway. There is a connection between the abstract theory of computation and constitutive physical theories. That is the physical version of the Church-Turing hypothesis which I have called the *Church-Turing principle* [6]:

Every finitely realizable physical system can be perfectly simulated by a universal model computing machine operating by finite means.

In the terminology of this paper a physical system $\mathfrak{S}$ "perfectly simulates" another system $\mathfrak{C}$ if $\mathfrak{S}$ can be programmed to be denotationally equivalent to $\mathfrak{C}$. That is, when part of $\mathfrak{S}$ is prepared in a suitable state in which a description of the dynamics of $\mathfrak{C}$ and a time interval T are encoded, another part of $\mathfrak{S}$ evolving for a given period, or until it signals that it has halted, is denotationally equivalent to $\mathfrak{C}$ evolving over the interval T. The "model" computing machine referred to need not be an actual physical object. Such a model is said to "exist" if its properties are idealized only in ways which the laws of physics do not prevent one from realizing arbitrarily closely in real objects. I have shown [6] that the Church-Turing principle is true of finite quantum systems in the absence of chronology violation by showing that there exists a *universal quantum computer* $\mathfrak{Q}$ which would perfectly simulate all finite quantum systems. One way of realizing $\mathfrak{Q}$ would be as a quantum computational network.

$\mathfrak{Q}$ is not universal with respect to chronology-violating quantum systems. The evolution in paradox 1, for instance, being nonlinear in the density operator, cannot be simulated by any chronology-respecting system including $\mathfrak{Q}$. On the other hand, the results of this paper make it plausible that a computer universal with respect to chronology-violating systems, let us call it $\mathfrak{Q}^{\pm}$, exists if the availability of negative-delay components is not forbidden by the laws of physics.

Note that negative-delay components, unlike all the other components of a computing machine, effectively have to be manufactured *in situ*. Considered as a computational resource, a finite chronology-violating region of spacetime is irreversibly depleted both by being used (because of its finite information-storage capacity) and by the mere passing of external time (because of its finite duration). Therefore $\mathfrak{Q}^{\pm}$, if it exists, must in effect contain a means of generating reliably, under program control, chronology-violating regions which could be used in computations. In other words, $\mathfrak{Q}^{\pm}$ exists only if it is possible to manufacture time machines. But the Church-Turing principle requires $\mathfrak{Q}^{\pm}$ to exist if and only if chronology violations (in the strong sense of closed paths for information) have a nonzero amplitude in the quantum state of the universe. Thus the Church-Turing principle implies the following conditional statement about the feasibility of time machines: *If there is a nonzero amplitude for a chronology violation to occur somewhere in spacetime, then it is possible in principle to manufacture time machines.*

It is perhaps worth stressing why the Church-Turing principle does not likewise imply, for instance, that quarks can be manufactured at will just because they exist somewhere in spacetime. Chronology violations are different because they allow new forms of computation. One can make something denotationally equivalent to a quark which is nevertheless not a quark. But to make something denotationally equivalent to a closed timelike line one needs a closed timelike line.

On the status of consistency conditions in physics

The results of this paper have been derived from a systematic application of what I have called the "kinematical consistency condition," namely that the state of any quantum system, as expressed by its density operator at a given time, is single valued.

I have given no argument in favor of this condition, nor do I believe that it is necessary to do so. As I have said, the quantities that are set equal in equations such as (2), (4), (6), (12), (15), (22), (24), and (33) are in each case simply two ways of describing the same thing. However, Friedman *et al.* [5], who impose the same condition, have argued that it is not a triviality but a substantive postulate. I wish to explain why I disagree with them.

They propose a "*principle of self-consistency*" which states that "*the only solutions to the laws of physics that can occur locally in the real Universe are those which are globally self-consistent.*" They argue that this principle is "not totally tautological" because a non-self-consistent "solution to the laws of physics" (i.e., a nonsolution) might be a *bone fide* solution under some new physical theory—in this case a theory which would give meaning to a "many-valued wave function." And they say that "if one is inclined from the outset to ignore or disregard the possibility of new physics, then one will regard self-consistency as a trivial principle."

That is a mistake. The requirement for self-consistency in scientific argument *is* a tautology, notwithstanding that is is always possible that future theories will assign meanings to propositions that are meaningless under the existing theory. One must be prepared to contemplate the falsity of any nontautological propostion; therefore, consider what it would mean to deny the "principle of self-consistency." An author whose paper is rejected by referees on the grounds that its purported result was self-contradictory could validly object that their criticism contained a tacit assumption of the "principle of self-consistency," which might be false. Of course it is true than any "contradiction" in the rejected paper could indeed have the same form as a valid implication of a future theory. In the present case, it is undeniably possible that "many-valued wave functions" could be a feature of the successor to quantum theory. Nevertheless it is a tautology, not needing to be postulated separately, that conclusions drawn from existing physics should follow consistently from the postulates of existing physics, and that conclusions drawn from any new theory should follow consistently from the new theory, which must be stated before conclusions can be drawn from it.

Conversely, the speculation that a future theory will allow "many-valued wave functions" does not contradict the "principle of self-consistency" of Friedman *et al.* For under such a theory multivalued wave functions "occurring locally" *would* be "globally self-consistent." Therefore the principle does not do the job that it was invented for, namely to rule out (for the sake of argument) such theories.

What Friedman *et al.* really wish to say is that their current work is based (as is mine) on the hypothesis that existing physical theories are true, or at least sufficiently so to describe the physical effects of chronology violation. We explore the consequences of that hypothesis without for a moment denying that it may turn out to be false. That is a matter for experiment. Having made that hypothesis, however, Friedman *et al.* are repeating themselves when they postulate separately that, for instance, wave functions are single valued.

It may seem pedantic to quarrel about the logical status of a proposition which, for whatever reason, all serious investigators are agreed upon. But I believe that this false classification can be misleading in several ways. First, it is related to the mistaken assumption that the "contradictions" of paradoxes 1–3 would, under classical physics or any other theory where they occurred, constitute proofs that closed timelike lines do not occur in nature. Second, one is led to analyze the acceptability or otherwise of chronology violations in logical rather than epistemological terms and thereby to miss what is truly unphysical about classical chronology violation, namely the nonevolutionary creation of knowledge in paradox 4, and to miss the fact that a substantive principle (the evolutionary principle) is required to rule out such processes. Third, it leads one to give excessive credence to the speculation that there might be "multivalued wave functions" in a future version of quantum theory, a speculation for which there is no motivation independent of the belief that there is a substantive "principle of consistency" which might be false and to which such a theory would be the alternative.

SUMMARY

I hope that I have persuaded the reader that most of the physical questions raised by the possibility of chronology violation are at root quantum computational questions. The conventional spacetime-geometrical methods of addressing them are therefore perverse: One first translates the questions into a language in which they cannot be cleanly expressed, and in which a host of unnecessary side issues must be addressed first, and in which it is extremely difficult to take quantum mechanics into account. Then one labors through considerable technical difficulties. And then one translates back. It is like using general relativity to prove the insolubility of the halting problem—no doubt it would be possible, but it would be neither efficient nor illuminating.

I have shown that the traditional "paradoxes" of chronology violation, whatever position one takes on their seriousness, do not occur at all under quantum mechanics. The real problem with closed timelike lines under classical physics is that they could be used to generate knowledge in a way that conflicts with the principles of the philosophy of science, specifically with the evolutionary principle. Supplementary data are, in general, required within chronology-violating regions, in addition to the initial data at the past boundary of spacetime, to fix a solution of the dynamical equations. I have postulated a maximum entropy rule to fix the supplementary data and I have conjectured that the evolutionary principle is generically satisfied by the application of that rule without any constraint being imposed on the initial data.

The physics near closed timelike lines is dominated by macroscopic quantum effects and has many novel features. The correspondence principle is violated. Pure states evolve into mixed states. The dynamical evolution is not unitary nor is it even the restriction to a subsystem of unitary evolution in a larger system. It is possible to "clone" quantum systems and to measure the state of a quantum system. The subjective probabilities of events can be different for different observers, even if all observers continue to exist throughout. "Asymmetric separation" between two observers is possible, whereby A may be separated from B even though B is not separated from A. Qualitatively new forms of computation are possible, and it is likely that there are improvements in the efficiency of existing forms.

All these effects are stable and do not require the maintenance of quantum coherence. They therefore apply to macroscopic systems. Rival versions of quantum theory give qualitatively different predictions about many of the effects. This would provide a crucial experimental test of the Everett interpretation against all others.

Global conservation laws continue to hold, but only at the level of expectation values even if locally they hold at the operator level. The second law of thermodynamics holds.

The Church-Turing principle implies that if there is a nonzero amplitude for a closed timelike line to occur somewhere in spacetime then it is possible in principle to manufacture time machines.

It is curious that the analysis of a physical situation which might well not occur should yield so many insights into quantum theory. But that is the nature of thought experiments. Perhaps we should also bear in mind that a frequently observed effect of time is to convert thought experiments into real ones.

ACKNOWLEDGMENTS

I wish to thank Dr. Charles Bennett, Dr. Artur Ekert, and Dr. Michael Lockwood for interesting discussions on the subject of this paper.

[1] E. H. Kronheimer and R. Penrose, Proc. Cambridge Philos. Soc. **63**, 481 (1967).
[2] B. Carter, Gen. Relativ. Gravit. **1**, 349 (1971).
[3] S. W. Hawking, and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge University Press, Cambridge, England, 1973), Sec. 6.4.
[4] F. J. Tipler, Ph.D thesis, University of Michigan, University Microfilms, Ann Arbor, Michigan, 1976.
[5] J. Friedman, M. S. Morris, I. D. Novikov, F. Echeverria, G. Klinkhammer, K. S. Thorne, and U. Yurtsever, Phys. Rev. D **42**, 6, 1915 (1990).
[6] D. Deutsch, Proc. R. Soc. (London) **A400**, 97 (1985).
[7] D. Deutsch, Proc. R. Soc. (London) **A425**, 73 (1989).
[8] M. Dummett, in *The Nature of Time,* edited by R. Flood and M. Lockwood (Basil Blackwell, Oxford, 1986).
[9] C. H. Bennett, IBM J. Res. Dev. **6**, 525 (1973).
[10] K. R. Popper, Proc. British Academy **46**, 3 (1960).
[11] K. R. Popper, *Objective Knowledge* (Oxford University Press, Oxford, England, 1972).
[12] In dynamics any change occurring over time in a physical system is called "evolution." In biology and epistemology the word refers only to changes that increase the adaptive complexity (a.k.a. knowledge) of the system by means of variation and selection. This is an unfortunte conflict of terminology, but I shall continue to use the same word to mean either thing, according to context.
[13] F. J. Tipler, Gen. Relativ. Gravit. **8**, 713 (1977).
[14] B. S. DeWitt, Phys. Rev. **160**, 1113 (1967).
[15] R. Penrose, in *General Relativity, an Einstein Centenary Survey,* edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, England, 1979).
[16] M. R. Garey, and D. S. Johnson, *Computers and Intractability* (Freeman, New York, 1979).
[17] I use the term "mixed state" to denote the information encoded in the density operator of an individual quantum system, where that operator does not have the dyadic form $|\psi\rangle\langle\psi|$. It is not intended to suggest a statistical ensemble.
[18] $|a\ b\rangle$ is shorthand for $|a\rangle|b\rangle$.
[19] This is similar in form but not in substance to the *method of maximum entropy* that is sometimes advocated in decision theory. In that case one hopes to guess a state that is in some subjective sense "most likely" to be close to the actual state of a system about which one has less than maximal information. In this case I am proposing a criterion that would objectively determine the actual state as a function of maximal initial data, whether or not anyone has information about either.
[20] In that the state $2^{-n}\hat{1}$ can be generated efficiently by a short *quantum* computer program (see Ref. [6]).
[21] The particles referred to here are the real ones being simulated by the network of Fig. 4. The number of carriers in the final state of the network is of course still 2.
[22] I have shown elsewhere [23,24] that the Everett interpretation, in which the quantum superposition principle remains unmodified, is in principle experimentally distinguishable from other "interpretations," all of which modify the superposition principle and should properly be called alternative versions of quantum theory.
[23] D. Deutsch, Int. J. Theor. Phys. **24**, 1 (1985).
[24] D. Deutsch, in *Quantum Concepts in Space and Time,* edited by C. J. Isham and R. Penrose (Oxford University Press, Oxford, England, 1986).
[25] It is debatable whether that is so under collapse "interpretations" when no observation has taken place. But it will certainly be so in paradox 3 if, as in the traditional version, it is an *observer* who does or does not travel back in time.
[26] S. W. Hawking, Phys. Rev. D **14**, 2460 (1976).
[27] C. H. Bennett, Int. J. Theor. Phys. **21**, 905 (1982).
[28] J. D. Barrow and F. J. Tipler, *The Anthropic Cosmological Principle* (Clarendon, Oxford, 1986).
[29] C. Sagan and W. I. Newman, Quart. J. R. Astron. Soc. **24**, 113 (1983).
[30] F. Echeverria, G. Klinkhammer, and K. S. Thorne, Phys. Rev. D **44**, 1077 (1991).
[31] W. K. Wootters and W. H. Zurek, Nature (London) **299**, 802 (1982).
[32] J. D. Bekenstein, Phys. Rev. D **23**, 287 (1981).
[33] In the ontology of classical physics there is no difference between saying that something *can* exist and saying that it *does* exist. Under quantum theory one can draw a distinction between things which have a unit or near-unit probability in the state of the Universe (or in a relative state under discussion) and those which merely have a nonzero probability—the former "exist" and the latter "can exist".

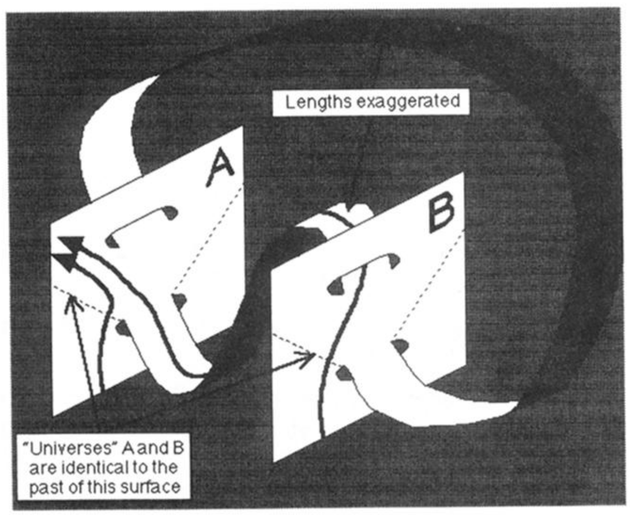

FIG. 6. Spacetime as perceived by observers in paradox 3.