Chapter 7
Information Loss

Gravity is perhaps not the only refinement that may guide us towards better models. Another interesting modification—though possibly related—might be of help. We shall now discuss *information loss* [9, 108].

7.1 Cogwheels with Information Loss

Let us return to the Cogwheel Model, discussed in Sect. 2.2. The most general automaton may have the property that two or more different initial states evolve into the same final state. For example, we may have the following evolution law involving 5 states:

$$(4) \rightarrow (5) \rightarrow (1) \rightarrow (2) \rightarrow (3) \rightarrow (1). \tag{7.1}$$

The diagram for this law is a generalization of Fig. 2.1, now shown in Fig. 7.1. We see that in this example state #3 and state #5 both evolve into state #1.

At first sight, one might imagine to choose the following operator as its evolution operator:

$$U(\delta t) \overset{?}{=} \begin{pmatrix} 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}. \tag{7.2}$$

However, since there are two states that transform into state #1 whereas there are none that transform into state #4, this matrix is not unitary, and it cannot be written as the exponent of $-i$ times an Hamiltonian.

One could think of making tiny modifications in the evolution operator (7.2), since only infinitesimal changes suffice to find some sort of (non Hermitian) Hamiltonian of which this then would be the exponent. This turns out not to be such a good idea. It is better to look at the physics of such models. Physically, of course, it
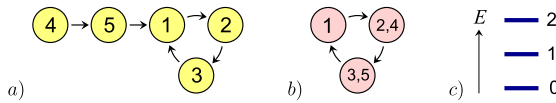
Fig. 7.1 **a** Simple 5-state automaton model with information loss. **b** Its three equivalence classes. **c** Its three energy levels

is easy to see what will happen. States #4 and #5 will only be realized once if ever. As soon as we are inside the cycle, we stay there. So it makes more sense simply to delete these two rather spurious states.

But then there is a problem: in practice, it might be quite difficult to decide which states are inside a closed cycle, and which states descend from a state with no past ("gardens of Eden"). Here, it is the sequence #4, #5. but in many more realistic examples, the gardens of Eden will be very far in the past, and difficult to trace. Our proposal is, instead, to introduce the concept of *information equivalence classes* (info-equivalence classes for short):

> *Two states (a) and (b) at time t_0 are equivalent if there exists a time $t_1 > t_0$ such that, at time t_1 , state (a) and (b) have evolved into the same state (c).*

This definition sends states #5 and #3 in our example into one equivalence class, and therefore also states #4 and #2 together form an equivalence class. Our example has just 3 equivalence classes, and these classes do evolve with a unitary evolution matrix, since, by construction, their evolution is time-reversible. Info-equivalence classes will show some resemblance with *gauge equivalence classes*, and they may well actually be quite large. Also, the concept of *locality* will be a bit more difficult to understand, since states that locally look quite different may nevertheless be in the same class. Of course, the original underlying classical model may still be completely local. Our pet example is Conway's *game of life* [39, 40]: an arbitrary configuration of ones and zeros arranged on a two-dimensional grid evolve according to some especially chosen evolution law, see Sect. 1.4. The law is not time-reversible, and information is lost at a big scale. Therefore, the equivalence classes are all very big, but the total number of equivalence classes is also quite large, and the model is physically non-trivial. An example of a more general model with information loss is sketched in Fig. 7.2. We see many equivalence classes that each may contain variable numbers of members.

Thus, we find how models with information loss may still be connected with quantum mechanics:

> *Each info-equivalence class corresponds to an element of the ontological basis of a quantum theory.*

Can information loss be helpful? Intuitively, the idea might seem to be attractive. Consider the measurement process, where bits of information that originally were properties of single particles, are turned into macroscopic observables. These may be considered as being later in time, and all mergers that are likely to happen must have taken place. In other words, the classical states are obviously represented
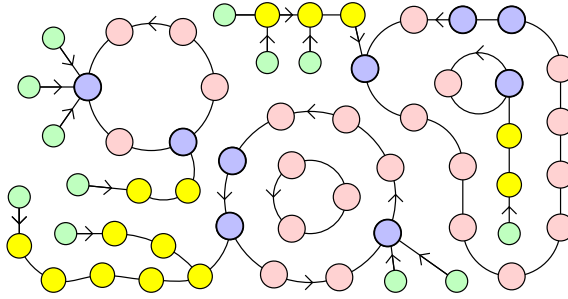
Fig. 7.2 Example of a more generic finite, deterministic, time non reversible model. *Largest* (*pink* and *blue*) *dots*: these also represent the equivalence classes. *Smallest* (*green*) *dots*: "gardens of Eden". *Heavier dots* (*blue*): equivalence classes that have "merge sites" among their members. The info-equivalence classes and the energy spectrum are as in Figs. 2.2 and 2.3

by the equivalence classes. However, when we were still dealing with individual qubits, the mergers have not yet taken place, and the equivalence classes may form very complex, in a sense "entangled", sets of states. Locality is then difficult to incorporate in the quantum description, so, in these models, it may be easier to expect some rather peculiar features regarding locality—perhaps just the thing we need.

What we also need is a better understanding of black holes. The idea that black holes, when emitting Hawking radiation, do still obey quantum unitarity, which means that the Hamiltonian is still Hermitian, is gaining in acceptance by researchers of quantum gravity, even among string theorists. On the other hand, the classical black hole is surrounded by a horizon from which nothing seems to be able to escape. Now, we may be able to reconcile these apparently conflicting notions: the black hole is an example of a system with massive amounts of information loss at the classical level, while the quantum mechanics of its micro-states is nevertheless unitary. The micro states are not the individual classical states, but merely the equivalence classes of classical states. According to the holographic principle, these classes are distributed across the horizon in such a way that we have one bit of information for each area segment of roughly the Planck length squared. We now interpret this by saying that all information passing through a horizon disappears, with the exception of one bit per unit horizon area.

We return to Hawking radiation in Sect. 9.4.

7.2 Time Reversibility of Theories with Information Loss

Now when we do quantum mechanics, there happens to be an elegant way to restore time reversibility. Let us start with the original evolution operator $U(\delta t)$, such as the one shown in Eq. (7.2). It has no inverse, but, instead of U^{-1}, we could use $U^{\dagger}$ as the operator that brings us back in time. What it really does is the following: the operator $U^{\dagger}(\delta t)$, when acting on an ontological state $|\text{ont}(t_0)\rangle$ at time t_0, gives us the additive quantum superposition of all states in the past of this state, at time

$t = t_0 - \delta t$. The norm is now not conserved: if there were N states in the past of a normalized state $|\text{ont}(t_0)\rangle$, the state produced at time $t_0 - \delta t$ now has norm $\sqrt{N}$. If the state $|\text{ont}(t_0)\rangle$ was a garden of Eden, then $U^\dagger |\text{ont}(t_0)\rangle = 0$.

Now remember that, whenever we do quantum mechanics, we have the freedom to switch to another basis by using unitary transformations. It so happens that with any ontological evolution operator U_1 that could be a generalization of Eq. (7.2), there exists a unitary matrix X with the property

$$U_1^\dagger X = X U_2, \tag{7.3}$$

where U_2 again describes an ontological evolution with information loss. This is not hard to prove. One sees right away that such a matrix X should exist by noting that $U_1^\dagger$ and U_2 can be brought in the same normal form. Apart from the opposite time ordering, U_1 and U_2 have the same equivalence classes.

Finding the unitary operator X is not quite so easy. We can show how to produce X in a very simple example. Suppose U_1 is a very simple $N \times N$ dimensional matrix D of the form

$$D = \begin{pmatrix} 1 & 1 & 1 & \cdots \\ 0 & 0 & 0 & \\ 0 & 0 & 0 & \\ \vdots & & & \end{pmatrix}, \qquad D^\dagger = \begin{pmatrix} 1 & 0 & 0 & \cdots \\ 1 & 0 & 0 & \\ 1 & 0 & 0 & \\ \vdots & & & \end{pmatrix}, \tag{7.4}$$

so D has N 1's on the first row, and 0's elsewhere. This simply tells us that D sends all states $|1\rangle, \ldots |N\rangle$ to the same state $|1\rangle$.

The construction of a matrix Y obeying

$$D^\dagger Y = Y D, \tag{7.5}$$

can be done explicitly. One finds

$$Y_{k\ell} = \frac{1}{\sqrt{N}} e^{2\pi i k\ell/N}. \tag{7.6}$$

Y is unitary and it satisfies Eq. (7.5), by inspection.

This result may come unexpected. Intuitively, one might think that information loss will make our models non-invariant under time reversal. Yet our quantum mechanical tool does allow us to invert such a model in time. A "quantum observer" in a model with information loss may well establish a perfect validity of symmetries such as CPT invariance. This is because, for a quantum observer, transformations with matrices X merely represent a transition to another orthonormal basis; the matrix Y is basically the discrete Fourier transform matrix. Note that merger states (see Fig. 7.2) transform into Gardens of Eden, and vice versa.

7.3 The Arrow of Time

One of the surprising things that came out of this research is a new view on the arrow of time. It has been a long standing mystery that the local laws of physics

appear to be perfectly time-reversible, while large-scale, classical physics is not at all time-reversible. Is this not a clash with the reduction principle? If large-scale physics can be deduced from small-scale physics, then how do we deduce the fact that the second law of thermodynamics dictates that entropy of a closed system can only increase and never decrease?

Most physicists are not really worried by this curious fact. In the past, this author always explained the 'arrow of time' by observing that, although the small-scale laws of Nature are time-reversible, the boundary conditions are not: the state of the universe was dictated at time $t = 0$, the Big Bang. The entropy of the initial state was very small, probably just zero. There cannot be any boundary condition at the Big Apocalypse, $t = t_\infty$. So, there *is* asymmetry in time and that is that. For some reason, some researchers are not content with such a simple answer.

We now have a more radical idea: the microscopic laws may not at all be time-reversible. The *classical* theory underlying quantum mechanics does not have to be, see Sect. 7.1. Then, in Sect. 7.2, we showed that, even if the classical equations feature information loss at a great scale—so that only tiny fractions of information are preserved—the emerging quantum mechanical laws continue to be exactly time-reversible, so that, as long as we adhere to a description of things in terms of Hilbert space, we cannot understand the source of time asymmetry.

In contrast, the *classical*, ontological states are very asymmetric in time, because, as we stated, these are directly linked to the underlying classical degrees of freedom.

All this might make information loss acceptable in theories underlying quantum theory. Note, furthermore, that our distinction of the *ontological states* should be kept, because classical states are ontological. Ontological states do *not* transform into ontological states under time reversal, since the transformation operators X and Y involve quantum superpositions. In contrast, templates *are* transformed into templates. This means automatically that *classical states* (see Sect. 4.2), are not invariant under time reversal. Indeed, they do not look invariant under time reversal, since classical states typically obey the rules of thermodynamics.

The quantum equations of our world are invariant (more precisely: covariant) under time reversal, but neither the *sub*-microscopical world, where the most fundamental laws of Nature reign, nor the classical world allow for time reversal.

Our introduction of information loss may have another advantage: two states can be in the same equivalence class even if we cannot follow the evolution very far back in time. In practice, one might suspect that the likelihood of two distinct states to actually be in one equivalence class, will diminish rapidly with time; these states will show more and more differences at different locations. This means that we expect physically relevant equivalence classes to be related by transformations that still look local at the particle scale that can presently be explored experimentally. This brings us to the observation that *local gauge equivalence classes* might actually be identified with information equivalence classes. It is still (far) beyond our present mathematical skills to investigate this possibility in more detail.

Finally, note that info-equivalence classes may induce a subtle kind of *apparent* non-locality in our effective quantum theory, a kind of non-locality that may help to accept the violation of Bell's theorem (Sect. 3.6).

The explicit models we studied so-far usually do not have information loss. This is because the mathematics will be a lot harder; we simply have not yet been able to use information loss in our more physically relevant examples.

7.4 Information Loss and Thermodynamics

There is yet another important novelty when we allow for information loss, in particular when it happens at a large scale (such as what we expect when black holes emerge, see above). Neither the operator U nor $U^\dagger$ are unitary now. In the example of Eq. (7.4), one finds

$$DD^\dagger = N|1\rangle\langle 1|, \qquad D^\dagger D = N|e\rangle\langle e|, \tag{7.7}$$

where $|e\rangle$ is the normalized state

$$|e\rangle = \tfrac{1}{\sqrt{N}}\big(|1\rangle + \cdots X|N\rangle\big) \tag{7.8}$$

(which shows that the matrix Y here must map the state $|1\rangle$ onto the state $|e\rangle$. Note, that D and $D^\dagger$ are in the same conjugacy class). Thus, during the evolution, the state $|1\rangle$ may become more probable, while the probabilities of all other states dwindle to zero. Some equivalence classes may gain lots of members this way, while others may stay quite small. In large systems, it is unlikely that the probability of a class vanishes altogether, so it might become possible to write the amplitudes as

$$e^{-iHt - \frac{1}{2}\beta E}, \tag{7.9}$$

where one might be tempted to interpret the quantity E as a (classical) energy, and $\frac{1}{2}\beta$ as an imaginary component of time. This aspect of our theory is still highly speculative. Allowing time to obtain complex values can be an important instrument to help us understand the reasons why energy has a lower bound.