Reversibility, irreversibility: restorability, non-restorability

1999 J. Phys. A: Math. Gen. 32 7581

(http://iopscience.iop.org/0305-4470/32/43/310)

J. Phys. A: Math. Gen. **32** (1999) 7581–7602. Printed in the UK PII: S0305-4470(99)00814-8

Reversibility, irreversibility: restorability, non-restorability

B Bernstein† and T Erber‡§

† Department of Chemical and Environmental Engineering and Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL 60616, USA
‡ Enrico Fermi Institute, University of Chicago, Chicago, IL 60637, USA

Received 12 January 1999

Abstract. Simple dynamical systems of point particles are irreversible if their motions cannot be retraced merely by reversing their velocity components. More complicated systems, such as those that exhibit steady hysteresis under the cyclic action of external influences, may be locally reversible, globally irreversible, and yet traverse an ordered set of recurrent states. In still more complex situations incorporating memory dependences, hysteresis effects are generally evolutionary or non-recurrent: nevertheless, in special circumstances, it may be possible to achieve a restoration of some prior states. An example based on the behaviour of an elastic–perfectly plastic torsion spring shows that such restorations require processes that are qualitatively more complicated than those associated with the original evolution of the systems. This asymmetry in the complexity of processes provides another means for assigning a direction to the arrow of time.

1. Reversibility, irreversibility, and Poincaré's non-existence theorem for 'one-way' evolution

A simple dynamical system, consisting of N independent point particles, is considered to be reversible if a reversal of all of its velocity components results in a retracing of the prior trajectories in inverse order. Formally, this reversion corresponds to an invariance of the equations of motion

$$\dot{p}_j = -\frac{\partial \mathcal{H}}{\partial q_j} + Q_j \qquad j = 1, \ldots, 3N \tag{1.1}$$

under the transformations $t \to -t$, $\dot{q}_j \to -\dot{q}_j$, and $p_j = -p_j$, where, as usual, q_j and p_j denote the generalized coordinates and momenta, $\mathcal{H}$ is the Hamiltonian, and $Q_j(q_i, \dot{q}_k, t)$ accounts for any additional external forces that may be present. Common exceptions arise when the kinetic energy contains terms linear in $\dot{q}$ due to Coriolis forces in rotating reference frames, or Q_j includes odd powers of $\dot{q}$ associated with the deflection of charges by magnetic fields. In these cases, (1.1) is not invariant under time reversal, and the corresponding dynamical systems are irreversible [1–4]. Of course, lack of retraceability of trajectories does not bar the recovery of prior configurations. If it is assumed that the dynamical systems are confined to bounded regions of phase space, Poincaré's recurrence theorem ensures that the trajectories through a point will return infinitely often to any neighbourhood of the point [5]. Moreover, with mild continuity restrictions on the dynamical flows, this result can be strengthened to show rigorously that it is impossible to construct any continuous function that varies monotonically along trajectories [2, 6, 7]. Consequently, Hamiltonian systems of

§ On leave from the Department of Physics, Illinois Institute of Technology, Chicago, IL 60616, USA.

0305-4470/99/437581+22$30.00

this type—even if they are irreversible in the preceding sense—are not sufficiently complex to generate any 'one-way' function that might be identified with a non-equilibrium entropy. Poincaré's non-existence result vitiates the deterministic form of Boltzmann's H-theorem [8], and shows that the second law of thermodynamics cannot be derived from the evolution of such smooth bounded Hamiltonian systems [2, 9, 10].

2.

2.1. Discontinuities and instabilities as sources of irreversibility

More realistic models of irreversible behaviour can be constructed by introducing additional physical interactions and relaxing the mathematical restrictions of completely smooth evolution: specifically, by weakening the presumptions of continuity to piece-wise continuity; allowing jump discontinuities; and admitting functions with disconnected domains of definition [11–13]. Mathematical generalizations of this type are associated with singularities ('catastrophes'), fractals [14], and also appear when smooth macroscopic variables such as strain, stress or magnetization are correlated with mesoscopic counterparts such as Portevin–LeChatelier discontinuities, acoustic emission or Barkhausen pulses [15]. Discontinuous response is actually quite common even in simple idealized systems whose physical states are determined by the competitive influences of internal and external forces. For example, quasi-static gradient systems such as arrays of N interacting pivoted dipoles (Ewing arrays [16]), or point charges confined to a sphere (Thomson's plum pudding atom [17]), generally have many locally stable equilibrium states whose multiplicity tends to increase exponentially with N. If a system of this type is initially prepared in one of these locally stable states, and an external 'force' Q_j—which may be a mechanical lattice distortion or an electric or magnetic field—is gradually applied, the system will generally accomodate by means of smooth perturbations of its equilibrium configurations. But as soon as at least one of the force components Q_j, acting on one or more of the individual dipoles or charges, reaches a critical threshold, the system will escape from the stability valley surrounding the initial state via a mountain pass—or Hessian instability point—and abruptly decant into one of the adjacent regions of stability. Since a reversal of the force components Q_j generally propels the system along an entirely different stability valley on the energy landscape, crossing a Hessian instability results in irreversible evolution. The precise connection between instabilities and discontinuities in gradient systems can be illustrated with the help of a simple example exhibiting hysteresis.

2.2. Hysteresis and irreversibility

Figure 1(a) is a schematic of an experimental arrangement consisting of two small cylindrical bar magnets—each with dipole moment μ and free to rotate in the horizontal x–y plane—mounted on low-friction bearings separated by a distance a. A uniform external magnetic field H^{ext}, generated by a set of Helmholtz coils, may be applied at an angle ψ with respect to the direction of the ordinate ($+y$) as shown. The total energy of this system, in emu units, is

$$U(\phi, H^{\text{ext}}, \psi) = -\frac{\mu^2}{2a^3}\{(1 - 3\cos(2\phi)) + [4h\cos(\phi - \psi)]\} \tag{2.1}$$

where $h \equiv H^{\text{ext}}/\mu a^{-3}$ is the dimensionless ratio of the external field and the natural scale of the internal, or interaction, field of the two dipoles (μ/a^3). Evidently, the first term in parentheses in (2.1) describes the interaction energy of the two dipoles, while the second term in square brackets represents the coupling to the external field ($\vec{\mu} \cdot \vec{H}$).

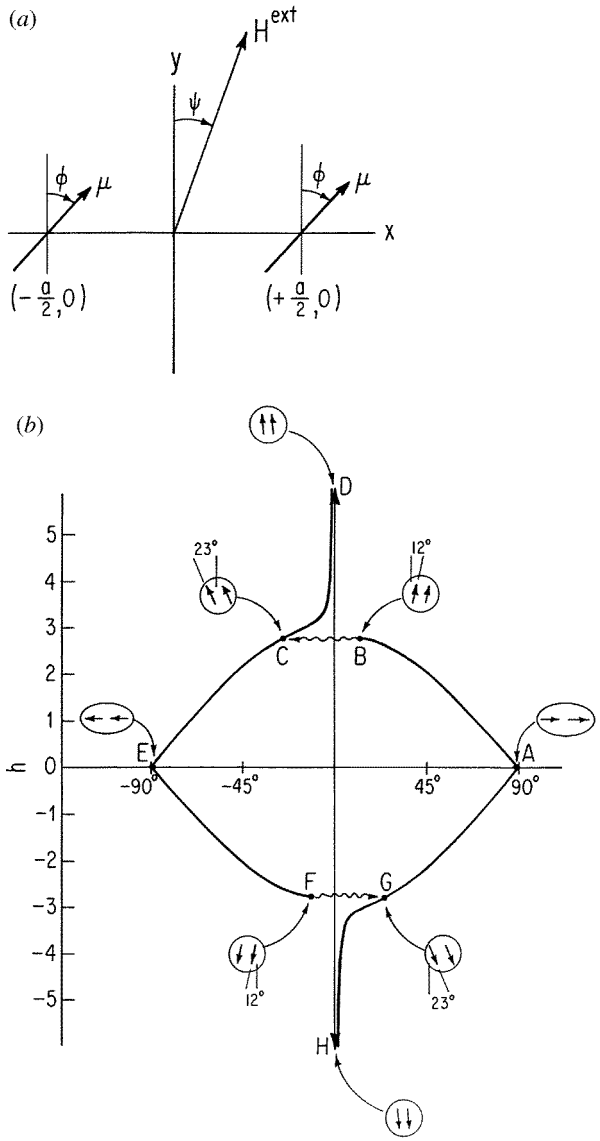

Figure 1. (*a*) Two pivoted magnetic dipoles interacting with an external field. This is a plan view of two cylindrical bar magnets, each with a magnetic moment $\mu \approx 14.5$ emu, length 0.89 cm, mounted on low-friction bearings separated by a distance $a = 5$ cm. The external magnetic field H^{ext} is generated by a set of Helmholtz coils (not shown). (*b*) Two-magnet hysteresis cycle with two irreversible jump discontinuities. This is a phase diagram of the stable trajectories determined by equation (2.1). The insets show the magnet orientations at various points of the cycle. In order to display the skew symmetry of the hysteresis, the ϕ-axis has been 'folded over', i.e., when $90^\circ < |\phi| < 180^\circ$, ϕ is replaced by $(180^\circ - |\phi|)\text{sgn}\,\phi$. (*c*) A hysteresis cycle without discontinuous jumps. This phase diagram shows the response of the two-magnet system to an external field H^{ext} restricted to varying in the vertical direction, i.e., $\psi = 0^\circ$ or 180° in (*a*). According to the discussion in the text, Q and P are quasi-reversible forks and also points of discontinuous magnetic susceptibility [18]. The abscissa is folded as in (*b*).

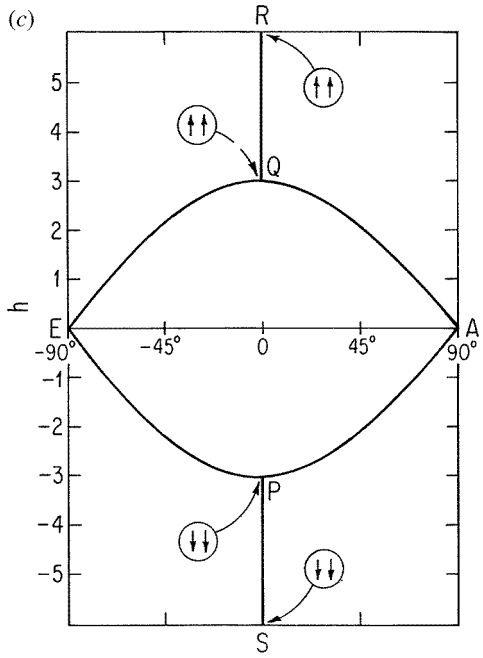

Figure 1. (Continued)

Solving for the extremals, namely solving

$$\frac{\partial U}{\partial \phi} = 0 \tag{2.2}$$

one obtains

$$3\sin(2\phi) - 2h\sin(\phi - \psi) = 0. \tag{2.3}$$

The stability criterion

$$\frac{\partial^2 U}{\partial \phi^2} > 0 \tag{2.4}$$

leads to the constraint

$$3\cos(2\phi) - h\cos(\phi - \psi) < 0. \tag{2.5}$$

Since non-parallel dipole configurations are unstable for all values of h, equations (2.1), (2.3) and (2.5) include all the essential physical features of this arrangement. The discontinuities implicit in the transcendental constraints in (2.3) and (2.5) are easier to recognize if H^{ext} is nearly aligned with the y-axis; i.e. $|\psi|$ is small. In this case, the trigonometric functions can be expanded to first order in ψ, and (2.3) and (2.5) may be approximated by

$$3\sin(2\phi) = 2h[\sin\phi - \psi\cos\phi] \tag{2.6}$$

and

$$3\cos(2\phi) < h[\cos\phi + \psi\sin\phi]. \tag{2.7}$$

With the additional restriction $\tan^2\phi > \psi^2$, these expressions can be combined into the simpler inequality

$$0 < \sin^2\phi + \psi\cot\phi \tag{2.8}$$

which is independent of the magnitude of h. It is then easy to verify that if H^{ext} is oriented as shown in figure 1(*a*), continuous and stable variations of the dipole directions are possible throughout the interval $0 < \phi \leqslant 90^\circ$. (The lower bound in this inequality follows from the exact condition $\tan(2\phi) > 2\tan(\phi - \psi)$ implied by (2.3) and (2.5).) However, if H^{ext} is rotated into the second quadrant of the x–y plane so that $\psi < 0$ in figure 1(*a*), the second term in (2.8) can become large and negative, and stability will cease as the magnitude of H^{ext} is increased. Specifically, if ψ is adjusted to the particular value $\psi = -0.5^\circ$, equation (2.8) implies that a transition to instability occurs when the dipoles have pivoted upwards to

$$\phi_{\text{disc}} \approx \sin^{-1}(|\psi|^{1/3}) \approx 12^\circ. \tag{2.9}$$

The magnitude of the associated normalized field can then also be inferred from (2.6) and (2.9):

$$h_{\text{disc}} \approx 3\cos^3\phi_{\text{disc}} \approx 2.8. \tag{2.10}$$

This instability threshold corresponds to the magnet orientations shown in inset 'B' in figure 1(*b*). The entire set of trajectories in this diagram—including the 'jump transitions' from B to C and F to G—form a hysteresis loop on the h–ϕ phase plane. The discontinuous behaviour of this system can be followed in detail by supposing that initially there is no external field ($H^{\text{ext}} = h = 0$), and that both magnets are aligned with the $+x$-axis ($\phi = 90^\circ$) as shown in inset A. The pole-reversed configuration at E corresponds to the other trivial solution of (2.3) when $h = 0$, i.e., $\phi = -90^\circ$. If an external field is gradually switched on—pointing in the fixed $\psi = -0.5^\circ$ direction—and increased continuously throughout the interval $0 \leqslant h < 2.8$, both magnets will respond by rotating through the angular range $90^\circ \geqslant \phi > 12^\circ$: This smooth evolution is represented by the trajectory joining A to B in figure 1(*b*). The discontinuous jump that occurs at B has a simple physical interpretation: since figure 1(*b*) is essentially a plane projection of the energy surface of the two-magnet system, the A $\rightarrow$ B trajectory actually corresponds to a valley bottom or trough that by virtue of the stability condition (2.4) is surrounded by 'higher ground'. At B, this trough terminates in a ledge that slopes downward to the other stability valley that connects E, C and D. The energy decrease that drives the B $\rightarrow$ C transition is given by (cf (2.1))

$$U(11.63^\circ, 2.819, -0.5^\circ) - U(-23.03^\circ, 2.819, -0.5^\circ) = \frac{\mu^2}{2a^3}\{-9.268 + 9.334\} \approx 0.033\frac{\mu^2}{a^3}. \tag{2.11}$$

The hysteresis circuit is completed by the F $\rightarrow$ G transition which flips the magnet orientations from $\phi \approx -168^\circ$ at F to $\phi \approx 157^\circ$ at G. In this case, the energy loss

$$U(-168.37^\circ, 2.819, 179.5^\circ) - U(156.97^\circ, 2.819, 179.5^\circ) \approx 0.033\frac{\mu^2}{a^3} \tag{2.12}$$

matches (2.11) because of the topographic symmetry of the energy surface. Overall energy conservation implies that the total energy dissipated during the irreversible jumps from B to C and F to G is equal to the energy gained along the reversible trajectories from G to B and C to F.

Experimentally, the slight misalignment between the axis of the Helmholtz coils and the perpendicular bisector of the line joining the magnet supports seems unimportant. However, since even this small deviation ($|\psi| \sim 0.5^\circ$) spoils the symmetry about the y-axis, its effects are quite noticeable. For instance, the jumps from B to C and F to G reorient both magnets by about 35°; and if the symmetry breaking angle ψ is switched from -0.5° to $+0.5^\circ$, the hysteresis loop in figure 1(*b*) will be traversed in a clockwise rather than an anti-clockwise direction. Figure 1(*c*) shows the curious hysteresis pattern that results when the field H^{ext} is exactly aligned with the y-axis. In this instance, (2.3) and (2.5) imply that the stable trajectories

consist of the (infinite) straight sections R–Q and P–S joined to the AQEP loop at the bifurcation points P and Q [18]. Since there are no discontinuities, the system is reversible and determinate everywhere except in the immediate vicinity of P and Q. The asymmetric indeterminacy of these saddle points is due to an energy degeneracy. If, for instance, the magnets are rotated along A $\rightarrow$ Q by increasing the field to the critical value $h \rightarrow 3$, their orientations will remain fixed in the vertical direction, $\phi = 0^\circ$, during all further field increases in the range $h > 3$. However, in the event that the field is subsequently decreased below $h \rightarrow 3-$, there is no unique response built into (2.1), (2.3) and (2.5): the magnets will continue to pivot either towards A ($\phi \rightarrow +90^\circ$) or E ($\phi \rightarrow -90^\circ$). A similar ambiguity occurs if h is increased precisely to 3, and then lowered again: this manoeuvre positions the system at Q, and then swivels the magnets back either towards A or E. If the field variations are increased to cover the entire range, $h < -3 \longleftrightarrow +3 < h$, extending beyond Q and P, the hysteresis response becomes erratic: sometimes the magnets will turn around completely through the hysteresis loop AQEP (cf section VI B of [19]), and at other times they will pivot through the half-cycles QES or QAS. This type of variable hysteresis without irreversible jumps, i.e. Barkhausen pulses, has been observed in iron whiskers (sections III D and IV D of [20]). Gallop's simple example is, of course, the antecedent of Hopf bifurcations [20a], and a forerunner of catastrophe theory [20b, 20c, 20d].

2.3. Smooth Hamiltonian systems and discontinuous gradient systems

Quasistatic magnet rotations in the simple model system of figure 1(*a*) can be included in the basic energy expression (2.1) by introducing a time dependence for the magnetic field, $h = h(t)$. The resulting equation of motion is then a special case of (1.1); and, according to the preceding discussion, if the (gradual) field variations are bounded by $|h| < 2.8$, the magnet response will be strictly reversible as in a smooth Hamiltonian system. The observed transition to irreversible behaviour at $|h| \geqslant 2.8$ depends on a number of physical effects that have been omitted from (2.1):

(1) Just past the instability points B and F in figure 1(*b*) the magnets are impelled across the downwards sloping energy surfaces towards C and G by gradient forces $-\nabla U$. This motion converts potential to rotational kinetic energy, $I\omega^2/2$. If nominal values ($\mu \sim 14.5$ G cm^3, $a \sim 5$ cm, $I \sim 0.043$ g cm^2) are assumed, the energy decreases (2.11) and (2.12) yield angular velocities of the order $\omega \sim 92^\circ$ s^{-1}, equivalent to rapid jumps from B to C etc in less than 0.4 s.

(2) The magnet motion also results in a coupling to several dissipative processes: (i) frictional resistance at the pivot supports; (ii) rotational spin-down due to dipole radiation; and (iii) magnetostrictive deformations of the ferromagnetic domains in the interior of the magnets. In practice, the magnetization changes and warming due to the magnetic work associated with (iii) are the most important dissipative processes [19]. The net result is that any instability jump away from B or F is followed by a rapid ($\lesssim 3$ s) re-equilibration at C or G.

(3) The mathematical description of the unstable equilibrium at B and F is deceptively simple. The torques exerted by the external field and the mutual magnet interactions are so evenly matched that both the first and second derivatives in (2.2) and (2.4) vanish. But it is precisely this cancellation of forces that shifts control of the physics from the energy expression in (2.1) to the dissipative processes mentioned above. Consequently, even in this elementary two-magnet device—which seems to be stripped down to irreducible simplicity—the root causes of irreversible behaviour are asymmetric boundary conditions (electromagnetic radiation to infinity), and the redistribution of energy into microscopic degrees of freedom (heating of the magnets).

3.

3.1. Irreversibility and restorability in complex systems

The incidence of discontinuities in the two-magnet model is not an isolated curiosity. Rather, it indicates the generic behaviour of more complicated gradient systems. It has already been mentioned in section 2.1 that N-component arrays of charges or magnets have approximately e^{cN} locally stable equilibrium states, where c is a configurational entropy [21]. In the usual energy landscape picture these states correspond to fixed localized minima. If changes are induced by external fields or lattice distortions, it is convenient to represent the resulting transformations by trajectories on 'extended' energy surfaces: these are simple generalizations of the diagram in figure 1(*b*). The key idea is to exhibit the dynamics by adding extra dimensions to the energy surfaces. This construction replaces the static picture of potential wells by sets of troughs or valleys that show explicitly how the equilibrium states are transformed. Figure 2 illustrates the situation for 16 cylindrical bar magnets supported by low-friction bearings, as in figure 1(*a*), and arranged in the form of a 4×4 square lattice. When the lattice spacing (a) is very much larger than the length of the individual magnets (ℓ), it is plausible that the observed equilibrium patterns are determined solely by dipole interactions. And, indeed, detailed computations, accounting for all 120 pairs of interactions, confirm that when $a \sim 140\ \text{mm} \gg 8.9\ \text{mm} = \ell$, the four configurations 4WSI–4WSIV shown in figure 2 are the stable equilibrium solutions of the 4×4 square dipole problem. The network of trajectories in figure 2 shows how these magnet patterns change as the lattice is compressed. Obviously, the dominant effect of bringing the magnets closer together is to break the dilatational invariance of the dipole interactions as higher multipolarities become more important. The net result is that when the lattice is contracted to $a \approx 11.6$ mm, the 16 magnets can appear in at least 36 distinct stable patterns.

Most of the 40 states in figure 2 are linked by irreversible transitions whose directions are indicated by the arrowheads on the wavy lines. These jump discontinuities occur at the ends of stability valleys in complete analogy with the B→C and F→G transitions in figure 1(*b*). A new feature of the circuit diagram in figure 2 is that it lacks closed hysteresis loops. This means that practically all of the dilations are irreversible as well as non-restorable. In simpler cases, such distinctions are unimportant. For instance, the hysteresis circuit in figure 1(*b*) is globally irreversible. Yet, any equilibrium state of the two-magnet system can be reached—or restored—starting from any other equilibrium state merely by applying a sequence of magnetic field variations. In contrast, repeated expansions and contractions of the 16-magnet lattice between $11.6\ \text{mm} \leqslant a \leqslant 36$ mm irretrievably decant the system into two tracks connected to the 4CS1 and 4UR2 configurations, irrespective of the initial state. Even the 'bottom trough' (the stability valley with the minimum energies) joining 4CS1 to 4WSI is not completely restorable. Any expansion of the system that crosses the quasi-reversible fork at $a \approx 53$ mm has roughly an 80% chance of remaining in the valley leading to 4WSI—the lowest energy dipole state for $a \gtrsim 140$ mm—and a 20% probability of diverting into the side channel leading to 4WSIII. Experiments and computer simulations confirm that both of the junctions at 53 mm and 79 mm in figure 2 are saddle points similar to the bifurcations of Q and P in figure 1(*c*). However, the topography of the junctions in figure 2 is asymmetric because these instabilities are due to multipole interactions with complicated angular variations [22].

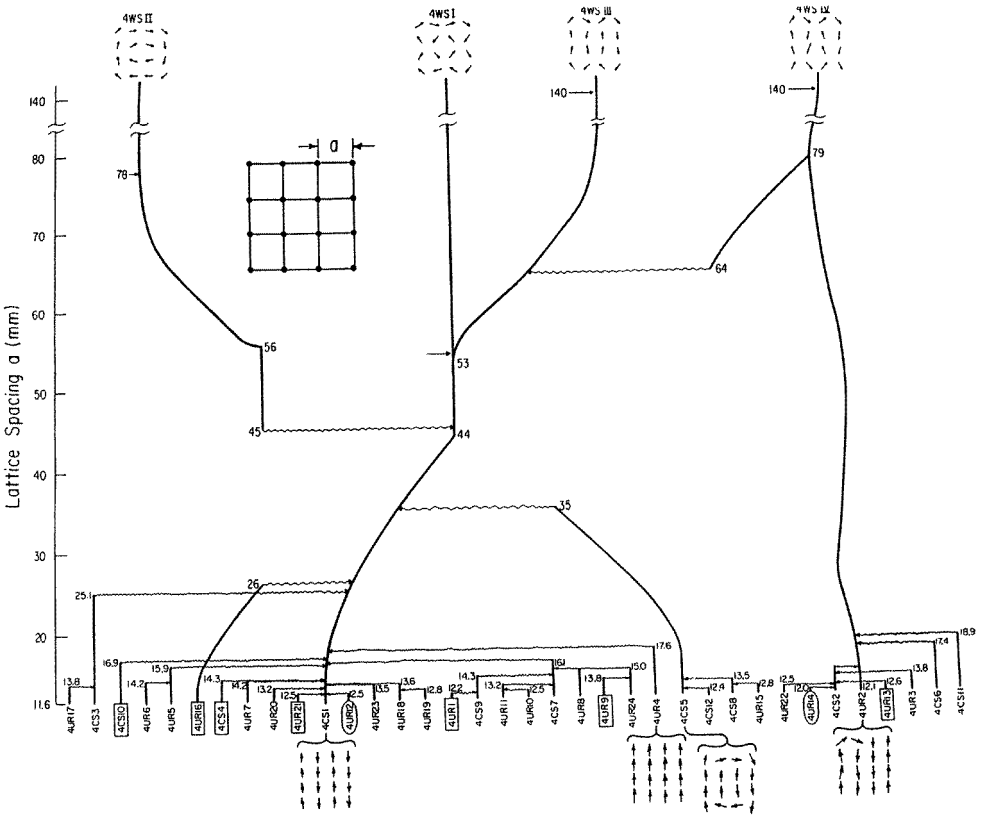

Figure 2. Hysteresis network for 16 pivoted magnets supported by an expandable 4×4 square lattice—shown in the inset. The solid straight lines represent magnet configurations that remain invariant during smooth lattice contractions or expansions; curved lines indicate gradual reorientations of the magnet patterns. The wavy lines capped by arrows denote irreversible jumps similar to those shown in figure 1(*b*).

3.2. Irreversibility and restorability in complex systems (continued)

The general significance of these results can be summarized as follows. Let

$$\mathcal{H}(\vartheta_1, \ldots, \vartheta_i, \ldots, \vartheta_N; h_c(t)) \tag{3.1}$$

represent an N-magnet Hamiltonian, where ϑ_i denotes the orientation of the ith magnet, and h_c is a hysteresis variable associated with lattice deformations or external fields. The physically realizable states are a subset of the extremals determined by the common intersection of the solutions of the equations (cf (2.2))

$$\frac{\partial \mathcal{H}}{\partial \vartheta_i} = 0 \qquad i = 1, \ldots, N. \tag{3.2}$$

Since for every value of h_c there are generally many solutions ($\sim O(e^{cN})$), it is convenient to write these in the form

$$\vartheta_i^{(\sigma)} = \vartheta_i^{(\sigma)}(h_c(t)) \tag{3.3}$$

where the index σ labels each distinct family of extremals. In quasi-static gradient systems the stable equilibrium states are local energy minima. These states are subsets of the extremal

solutions (3.3) whose associated Hessian matrices

$$H_e[\vartheta_i^{(\sigma)}(h_c(t))] = \left(\frac{\partial^2 \mathcal{H}}{\partial \vartheta_k^{(\sigma)} \partial \vartheta_j^{(\sigma)}}\right) \qquad k, j = 1, \ldots, N \tag{3.4}$$

when evaluated at $\vartheta_i^{(\sigma)}$ are positive definite (cf (2.4)). A necessary and sufficient condition for a real symmetric $N \times N$ matrix such as H_e to be positive definite is that all of its eigenvalues be positive. This property may be expressed in the compact form

$$H_e[\vartheta_i^{(\sigma)}(h_c(t))] \succ 0 \tag{3.5}$$

which is an intuitive shorthand for sufficient conditions ensuring local stability on the extremal set $\vartheta_i^{(\sigma)}(h_c(t))$. Some exceptional cases are discussed in [23, 24]. Transitions to instability generally occur whenever the Hessian becomes singular; that is when the determinant of the matrix in (3.4) vanishes

$$\det |H_e[\vartheta_i^{(\sigma)}(h_c(t))]| = 0. \tag{3.6}$$

Finally, if the Hessian has at least one negative eigenvalue—indicating a downward slope on the energy surface—it is natural to use the notation

$$H_e[\vartheta_i^{(\sigma)}(h_c(t))] \prec 0 \tag{3.7}$$

as a shorthand for sufficient conditions that the extremal $\vartheta_i^{(\sigma)}(h_c(t))$ is unstable.

Reversibility in these N-magnet systems simply means that there are extremals $\vartheta_i^{(\sigma)}(h_c(t))$ satisfying the stability criterion (3.5) for some continuous ranges of the variable $h_c(t)$. Within these ranges, inverting the hysteresis sequences—i.e., reversing the quasi-static 'velocities' $\mathrm{d}h_c/\mathrm{d}t$—results in retracing of the prior magnet configurations in complete analogy with the behaviour of the reversible Hamiltonian systems in section 1. The limits of the reversible ranges are determined by the Hessian singularity condition (3.6). Specifically, if $\det |H_e|$ vanishes when the hysteresis variable h_c on the σth extremal reaches the particular value $h_c(t_s)$ at time t_s, then the system becomes unstable. In cases where the further evolution $h_c(t)$, $t > t_s$ prods the system into an unstable region $H_e \prec 0$ (3.7), continuations along the σ-extremal $\vartheta_i^{(\sigma)}$ are impossible, and the system must 'roll down' the energy surface into another (not necessarily unique!) locally stable extremal $\vartheta_i^{(\tau)}$ where $H_e[\vartheta_i^{(\tau)}(h_c(t))] \succ 0$ for $t \geqslant t_s$. Since 'rolling' is physically equivalent to a reorientation of the magnets there will be at least one value of i where $\vartheta_i^{(\sigma)}(h_c(t_s)) \neq \vartheta_i^{(\tau)}(h_c(t_s))$: such a finite gap between instability and re-equilibration corresponds precisely to the jump discontinuities illustrated in figures 1(*b*) and 2. Alternatively, several stable extremals may intersect at a singular configuration ϑ_i^{sp} where $\det |H_e[\vartheta_i^{sp}]| = 0$. In this situation the system can pass through the ϑ_i^{sp} pattern without any discontinuous changes. However, as illustrated by the junctions at Q and P in figure 1(*c*), and at 53 mm and 79 mm in figure 2, repeated traversals of these furcations do not lead to repeatable results, and are therefore also irreversible. The precise classification of the manifold branchings associated with Hessian zeros is a prime concern of catastrophe theory [20c, 20d]: these references are also a rich source of further examples of systems exhibiting singularities and hysteresis.

In extremely simple systems *irreversibility* does not exclude *restorability*. This assertion nominally applies to the hysteresis cycle in figure 1(*b*) where any stable magnet pattern can be reset by field variations even if these include irreversible jumps. But the claim of 'making things new again' depends on the level of abstraction. The two-magnet hysteresis cycle is actually not retraced exactly if the energy losses accompanying re-equilibration are taken into account. According to equations (2.11) and (2.12), every cycle dissipates about 0.1 ergs into the internal degrees of freedom, resulting in a $10^{-8}\,^{\circ}$C temperature rise in each magnet. Prolonged hysteresis can amplify such minute ageing effects until the entire system is transformed (e.g.

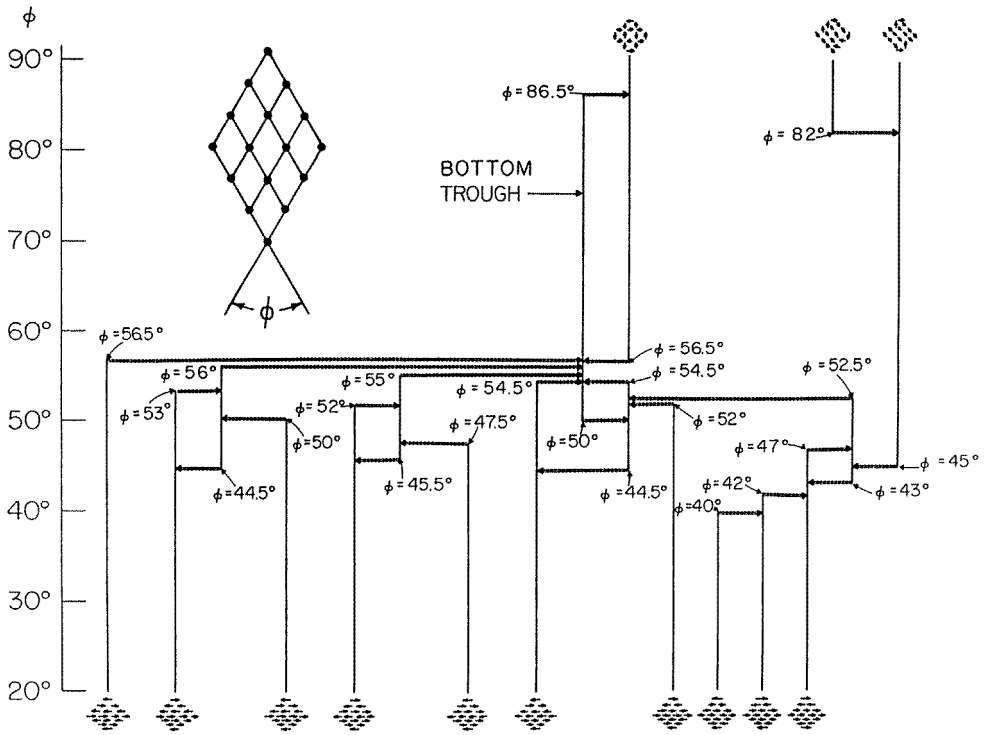

Figure 3. Hysteresis network for 16 pivoted magnets supported by a 4×4 trapezoidal hinged linkage—shown in the inset. In contrast to figure 2, where the hysteresis variable is the lattice spacing a, the complex hysteresis circuits in this diagram are generated by opening and closing the linkage angle ϕ.

by heating beyond the Curie temperature where all ferromagnetic properties are completely quenched). An important practical example is the cumulation of material defects in stress–strain hysteresis leading to catastrophic fatigue failure.

Restoring the state of a system after an irreversible transition is only feasible if there is at least another way 'back'. In the lattice dilation example in figure 2, such reversions are impossible for most states because only a few return paths are accessible when the lattice is contracted. However, when there are more complex interlaced hysteresis circuits—as illustrated in figure 3—adroit manipulation of the hysteresis variables can thread a path backwards through the maze and restore some initial states. Restorations can also be achieved by more drastic means. Suppose that the 16-magnet array in figure 2 were initially prepared in the ferromagnetic (4UR4) pattern shown in the inset, and that subsequently the lattice was expanded beyond 140 mm. The outcome would be either one of the irreversible transformations 4UR4 $\rightarrow$ 4WSI or 4UR4 $\rightarrow$ 4WSIII. Now, it is entirely feasible to recreate the ferromagnetic pattern at a spacing of 140 mm by applying a strong external magnetic field parallel to the lattice. As long as this field is maintained, the ferromagnetic pattern is essentially frozen in place, and the lattice can be contracted again to a spacing of 11.6 mm. The final step is to quench the external field. The net result of these double hysteresis variations is of course a (nearly perfect) restoration of the initial 4UR4 state.

All of these examples suggest that in systems of this type (e.g. (3.1)) there are pervasive correlations between increasing complexity and difficulties in finding return routes through

the hysteresis networks. Further, since these return routes generally require complicated manipulations of the hysteresis variables, it is evident that there are basic asymmetries between the initial evolution of states and their restoration. This trend parallels the second law of thermodynamics, but operates at a different level: whereas the standard connections between irreversibility and the increase of entropy ultimately depend on the *enumeration of states*, here the distinction between evolution and restoration depends on the relative *complexity of processes*. In plain terms, this implies that the arrow of time points in the direction of the simplest process! It is plausible that this conjecture applies to any system described by a Hamiltonian (3.1) that is periodic in each of the configuration coordinates ϑ_i. In these cases, $\mathcal{H}(\vartheta_1, \ldots, \vartheta_N; h_c(t))$ is a real-valued function on a closed manifold equivalent to the N-torus and topological estimates are available that set lower bounds on the number of extremal solutions [25]. In particular, Morse theory implies that the number of extremals $\vartheta_i^{(\sigma)}(h_c(t))$ in (3.3) exceeds $\sigma > \exp(0.7N)$ [19]: This bound is consistent with the observation that the number of locally stable states of N-component Ewing arrays is $\mathrm{O}(\exp(0.2N))$. Since the number of saddle points on an energy surface is comparable to the number of local minima, the set of Hessian singularities (3.6) also increases exponentially with N. Perturbations of such large systems are bound to generate hysteresis networks of great complexity.

4.

4.1. Restoring Humpty Dumpty?

The notion that time's arrow is an emergent property of complex hysteresis systems is also supported by probabilistic arguments. Combinatorial results [21] and studies of random Hamiltonians [26, 27] confirm the intuitive expectation that N-component systems have complex energy surfaces when $N \gg 1$. The corresponding number of Hessian singularities (3.6) can be estimated from the zeros of random matrix polynomials [28]. And, finally, the general characteristics of the associated hysteresis circuits can be inferred from Ramsey theory [29] and the probabilistic geometry of networks [30]. Since all of these results—including the topological bounds for the $N-$torus in section 3.2—remain valid for arbitrarily large values of N, the asymmetries between evolution and restoration also extend to macroscopic hysteresis systems. The tricky problem of restoring the shape of a bent metal paper clip is a familiar example. And, in fact, many other practical situations involving fatigue, ageing and breakdown are constrained by underlying limits of restorability. The general problem is the following: given a set of initial conditions and a sequence of irreversible transformations, just how far can a system be pushed before it reaches a limit of irreparable damage? It is important to realize that 'limit' in this context only has a meaning with respect to *particular classes of restorative processes*. Even in the simple magnetic models of section 3.2 there are cases where restorations are impossible using only lattice deformations, but other means relying on field variations can still bring the systems back to their initial states. More realistic illustrations of such asymmetric processes can be adapted from the theory of plasticity.

4.2. Restoring an elastic–plastic torsion spring

Figure 4 shows a system consisting of a rigid rod of length ℓ attached to a horizontal support by means of a frictionless pivot and an elastic–perfectly plastic torsion spring. As indicated, the angle between the rod and the vertical is Θ. A vertical load P is applied at the end of the rod, with the convention that the positive direction is downward. This system has the following attributes:

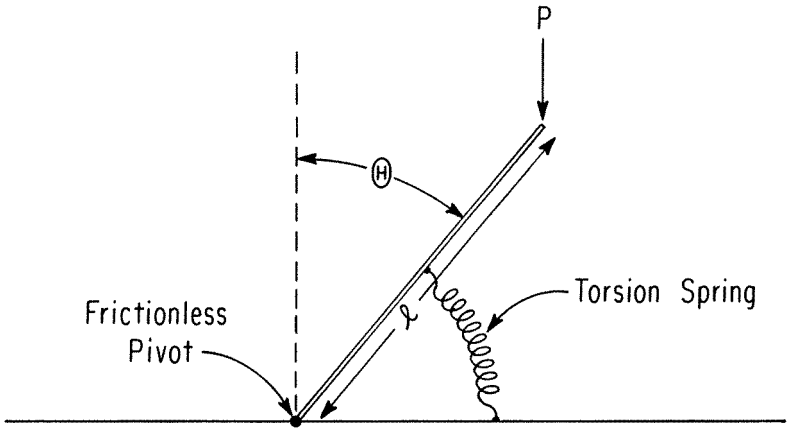

Figure 4. A plastic system that is irreversible and restorable. A rigid rod of length ℓ is attached to a horizontal support by means of a frictionless pivot. The plastic element is provided by a torsion spring connecting the rod and the support.

(i) It can be deformed into another state by a given set of forces.
(ii) Simple variants of this same set of forces cannot restore the original state.
(iii) However, a more general system of forces *can* restore the original state.

It is convenient to begin by describing the properties of the elastic–perfectly plastic torsion spring.

4.2.1. Constitutive equations. In addition to the angle Θ, it is necessary to introduce three other quantities—τ, the torque; $\mu > 0$, the elastic modulus of the spring†; and $k > 0$, the yield stress of the spring. Both μ and k are material constants. The relation between Θ and τ, i.e. the constitutive relation, is history dependent and is that of classical elastic–perfectly plastic behaviour. However, the present description may not be quite the usual one.

Instead of a single elastic relation between Θ and τ, we shall consider an infinite multiplicity of such relations, one for each value of an angle α, namely

$$\tau = \mu \cdot (\Theta - \alpha). \tag{4.1}$$

If (4.1) holds for a constant value of α, we shall say that the spring is in the *elastic regime* α. This assertion will continue to be valid as long as (4.1) does not give a stress exceeding the yield stress, i.e., for all torques $|\tau| < k$. However, once the stress given by (4.1) exceeds this bound, or $|\tau| > k$, then the spring has left the elastic regime α. In order to describe what comes next, we shall assume that Θ varies by increasing or decreasing in a piecewise manner, and shall consider the particular situation where the spring, starting in the elastic regime $\alpha \equiv \alpha_0$, experiences a monotonic increase of the difference $|\Theta - \alpha_0|$. Increasing this value corresponds to a *loading*.

In such a situation, beginning at the point where $\mu \cdot |\Theta - \alpha_0|$ reaches the value k, the spring is in a plastic regime; that is, it experiences *yield* and $|\tau|$ will remain fixed at the value k as long as $|\Theta - \alpha_0|$ is non-decreasing. An alternate way of describing this behaviour is the following: the relation (4.1) is always obeyed, but in an elastic regime α remains at a constant value, say possibly α_0, whereas in a plastic regime α varies so as to keep $|\Theta - \alpha|$ fixed at k/μ. Suppose now that the spring is in the plastic regime, and that α has reached a value, say $\alpha_1 > \alpha_0$,

† Not to be confused with the magnetic moments considered in section 2.2.

while loading. If this is followed by an unloading, the value of $|\Theta - \alpha_1|$, and consequently also the value of $|\Theta - \alpha_0|$, will decrease. As a result, the spring will be in the elastic regime α_1, and will remain therein as long as $|\Theta - \alpha_1|$ does not exceed k/μ. These variations fairly well describe the constitutive behaviour of the spring, except for one addtional qualification: For purposes of the present discussion, it will be assumed that plastic yield occurs well before $|\Theta - \alpha_0|$ reaches the value $\pi/2$.

Finally, we note that the behaviour of the spring can be modelled by the differential equation

$$\frac{\mathrm{d}\tau}{\mathrm{d}\Theta} = \mu \cdot \Big[H_\nu(\tau + k) - H_\nu(\tau - k) \Big] \tag{4.2a}$$

where H_ν is the Heaviside unit step function

$$H_\nu(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ 1 & \text{if} \quad x \geqslant 0. \end{cases} \tag{4.2b}$$

4.2.2. The initial elastic response of the system. According to figure 4, the torque due to the vertical (positive downward) load P is

$$\tau = P\ell \sin\Theta. \tag{4.3}$$

Suppose, for the sake of simplicity, that the initial value of Θ is adjusted to $\Theta = 0$. We shall argue that if all loads are restricted to the vertical direction, then the system permits Θ to be changed from zero to a value where the spring yields. However, once plastic yield has occurred, vertical loadings cannot restore the spring and lever arrangement to its initial configuration with $\Theta = 0$. Nevertheless, combinations of vertical and horizontal loadings can reset the system to $\Theta = 0$.

In the elastic regime, α, the system responds to a given load P by finding a stationary minimum value of the equivalent potential

$$W = \frac{\mu}{2}(\Theta - \alpha)^2 + P\ell\cos\Theta \tag{4.4}$$

or, explicitly

$$\mu \cdot (\Theta - \alpha) - P\ell \sin\Theta = 0. \tag{4.5}$$

These relations are the exact analogues of the corresponding expressions (2.1) and (2.3) that describe the magnetic model in figure 1(*a*). In the initial state, $\alpha = 0$, so that (4.5) simplifies to

$$\mu\Theta - P\ell\sin\Theta = 0 \tag{4.6}$$

which has the trivial solution $\Theta = 0$. If other solutions with $\Theta \neq 0$ exist, then evidently they must satisfy

$$\frac{\sin\Theta}{\Theta} = \frac{\mu}{P\ell} \tag{4.7}$$

which is possible only if $P > \mu/\ell$.

As usual, when there are multiple extremals, the physically relevant solutions can be distinguished by stability criteria such as (3.5). In the present case, equation (4.4) yields

$$\frac{\mathrm{d}^2 W}{\mathrm{d}\Theta^2} = \mu - P\ell\cos\Theta \tag{4.8}$$

whose sign—in analogy with (2.4)—determines whether W has a maximum or minimum. We note first that if $P\ell > \mu$, then (4.8) implies that $\mathrm{d}^2W/\mathrm{d}\Theta^2 < 0$ at $\Theta = 0$; and therefore W has

an unstable maximum at $\Theta = 0$. On the other hand, if $\Theta \neq 0$ and $P\ell > \mu$, then combining (4.6) and (4.7) with (4.8) gives

$$\frac{\mathrm{d}^2 W}{\mathrm{d}\Theta^2} = P\ell \left[\frac{\sin\Theta}{\Theta} - \cos\Theta \right] \tag{4.9}$$

which is positive for $-\pi < \Theta < \pi$, an interval exceeding the range in which the spring initially remains elastic. This assertion follows by noting that $\sin\Theta - \Theta\cos\Theta$ vanishes at $\Theta = 0$, and has a positive derivative, $\Theta\sin\Theta$, for all $|\Theta| < \pi$. Consequently, the stable solutions are determined by (4.7) when $\Theta \neq 0$. Indeed, when loaded, the spring is equally likely to bend either to the right or to the left. To be specific, we shall assume that it bends to the right, i.e. that $\Theta > 0$.

4.2.3. Non-restorability after plastic yield. Suppose that the spring has undergone plastic yield so that $\alpha > 0$. Either $\Theta = 0$ is now within the elastic regime α or it is not. If so, i.e. if $\alpha \leqslant k/\mu$, then (4.5) implies that $\alpha = 0$, which we have just assumed is not true. Worse, even if $\Theta > 0$, equation (4.5) shows that

$$P = \mu \frac{\Theta - \alpha}{\ell \sin\Theta} \to \infty \qquad \text{as} \quad \Theta \to 0. \tag{4.10}$$

So it is never possible to reach $\Theta = 0$ with a finite load P. But even if P could be increased indefinitely in such a way that (4.5) continued to be valid, the magnitude of α would remain at a non-vanishing value, $\alpha > 0$. As a consequence, any reversal in the sign of P would only have the effect of producing variations in the elastic regime for this positive α value, and therefore the system could never be returned to its initial state.

If $\Theta = 0$ is not in the elastic regime α, that is if $\alpha > k/\mu$, then in order to return to $\Theta = 0$, one must do so through plastic deformations. In this case, the value $\Theta = 0$ could be reached as α changes with the fixed constraint $|\Theta - \alpha| = k/\mu$. Therefore, if the initial condition $\Theta = 0$ is ever restored in the elastic regime, the corresponding α value would remain at $\alpha = k/\mu$; and we can argue as before that any reversals of the sign of P would leave the system in an elastic regime with $\alpha \neq 0$. Moreover, since $\mu|\Theta - \alpha| = k$, equation (4.5) requires that

$$P\ell \sin\Theta = k \tag{4.11}$$

which shows that P would have to go to infinity as Θ approached zero. In any case, it is not possible to reset α to zero, and therefore to restore the system to its original state. For completeness sake, we note that these results are not restricted to the special choice of the initial condition $\Theta = 0$.

4.2.4. Restoration after plastic yield. All of the preceding arguments show that the torsion spring device in figure 4 is irreversible and non-restorable with respect to a restricted set of load variations. However, if the vertical force P is augmented by horizontal load components, it is possible to adjust the torque τ in an arbitrary manner. In particular, it is then feasible to vary Θ so that it is brought to a value of $-k/\mu$ during yield: at this point, since $|\Theta - \alpha|$ remains fixed at k/μ during yield, α becomes zero again. Continuing on through the elastic regime to $\Theta = 0$ results in a complete restoration of the system to its initial state.

5. Four different kinds of time

The conventional time variable 't' initially appeared in the Hamiltonian equations of motion (1.1), and later was reintroduced in sections 2.3 and 3.2 to describe changes in the hysteresis

coordinates associated with lattice distortions and magnetic field variations. In those contexts, the direction of the arrow of time is indicated simply by the algebraic notation '$+t$' or '$-t$'. Evidently, there is a considerable gap—if not a total disconnection—between this standard notion of time as a parameter and the suggestion that time and its arrows are emergent properties of the transformations of complex systems. But these disparities are actually rooted in a lengthy history: the extensive literature concerned with 'time' includes many qualitatively different concepts regarding its duration and intrinsic nature, e.g. [31–50]. By linking together just four different kinds of time from this large inventory—*technical* time; *ordinal* time; *index* time; and *process* time—it will become apparent that there is in fact a reasonable progression leading from the usual time read off clocks and calendars to detecting the subtle alignments of temporal arrows embedded in complex processes.

5.1. Technical time (NIST-7 and UTC)

Practical applications of dynamics, such as Hamilton's equations of motion (1.1), depend on relating the formal coordinates, q_i and t, to specific physical frames of reference. The most widely used space–time grids are descendants of an 1884 treaty establishing Greenwich Mean Time (GMT) as the terrestrial standard, and designating the Greenwich Observatory as an intercept of the prime meridian. As early as 1902, Poincaré cautioned that the proliferation of standards of this kind could eventually blur the distinctions between conventions and 'physical reality' [51, 52]. He insisted that all conventions, despite their practical necessity, were in essence just artificial and largely arbitrary constructs. However, as shown later by Einstein [53] and Kretschmann [54], this extreme position could lead to sterile assertions such as '... covariance is merely a criterion concerning the mathematical formulation of physical laws and has no bearing whatever on their actual content'. In fact, relativistic corrections are indispensable for precision time keeping. Coordinated Universal Time (UTC), the current successor to GMT, depends on the synchronous operation of about 254 clocks distributed among 60 standard laboratories and 24 orbiting satellites of the Global Positioning System (GPS) [55, 56]. Since the present accuracy of UTC is equivalent to ± 1 s in 10^7 years, it is necessary to allow for at least 9 different time shifts due to special and general relativistic effects. For instance, the slowing of moving clocks implies that the GPS clocks lag behind an ideal clock situated at mean sea level (the geoid) by about 7.11 μs per day; and NIST-7, the current primary standard for the US, runs fast by about 15.55 ns per day relative to this ideal clock due to a difference in gravitational potential. The fusion of physics and conventions was carried still further by Schrödinger, who showed that the joint constraints of relativity and quantum mechanics placed additional limits on the accurate layout of coordinate grids [57]. Nevertheless, quantum mechanics in the form of atomic clocks has turned out to be a help rather than a hindrance [58].

In general, clocks may contain irreversible components such as escapements. But in all cases, it is essential that every 'tick' or clock cycle restore an initial state that ensures that the duration of the next cycle is an exact repetition. In this respect, ideal clocks are devices with no memory. Secular drifts of periodicity due to wear and ageing can occur only in complex systems with sufficiently many degrees of freedom to support evolutionary hysteresis, cf section 3.2. The superior stability of atomic clocks utilizing linear ion traps is partly due to the fact that the individual atoms cannot 'run down' because there are no smaller degrees of freedom.

5.2. Ordinal time in dynamical systems

The shift away from clock time to more general measures of evolution can be illustrated with the help of dynamical systems theory (e.g. [59]). In this formalism the evolution of a system through a set of states, $x_0, x_1, \ldots, x_j$, is described by the iteration of a mapping, i.e.

$$\begin{array}{ccccccccc} x_0 & \to & x_1 & \to & x_2 & \to & \cdots & \to & x_j \\ & & \updownarrow & & \updownarrow & & & & \updownarrow \\ x_0 & \to & f(x_0) & \to & f^2(x_0) & \to & \cdots & \to & f^j(x_0) \end{array} \tag{5.1}$$

where $f^2(x_0) = f(f(x_0)$, etc, or, equivalently

$$f^2 = f \circ f \tag{5.2}$$

where '$\circ$' denotes composition. In analogy with the Γ-space of classical statistical mechanics, the x_j specify the complete configuration of the system. For simplicity, it is convenient to assume that this configuration space $\mathcal{M}$ is an open subset of m-dimensional Euclidean space $\mathbb{R}^m$. An array of the form (5.1) is usually referred to as a 'discrete time' dynamical system because of the plausible identification of the iteration index with multiples of a fixed time interval. If f is a diffeomorphism, i.e., both f and the inverse function f^{-1} are continuously differentiable, then the dynamics described by (5.1) is sufficiently regular so that the sequence linking x_0 to x_j can be retraced in inverse order by repeated applications of f^{-1} to x_j. However, since the two functions f and f^{-1} are generally not related by reflections in the directions of any forces or momenta, and may be quite different in structure and complexity, retracing the states in (5.1) in inverse order does not correspond to time reversal in ordinary Hamiltonian dynamics. Rather, the two-step process $x_0 \to f(x_0) = x_1$, followed by $f^{-1}(x_1) \to x_0$, or concisely

$$f^{-1} \circ f = j \tag{5.3}$$

where $j(x) = x$ is the identity function, is the mathematical description of the *restoration* of a state.

Similar issues of interpretation occur in dynamical systems with continuous time. For instance, the logistic map

$$g(x) = 2x(1-x) \qquad 0 \leqslant x \leqslant 1 \tag{5.4}$$

can be embedded in an infinite family of functions of the form

$$g^t(x) = \tfrac{1}{2}\{1 - |1-2x|^{2^t}\} \tag{5.5}$$

where $-\infty < t < +\infty$ [60]. For integer values $t = 1, 2, \ldots, n$, this expression generates the iterates of the logistic map $g(x)$. Furthermore, (5.5) satisfies the semi-group property

$$g^{t+s} = g^t \circ g^s \tag{5.6}$$

for all real s and t. In particular, therefore, if $n \gg 1$ and $\ln|1-2x|$ is bounded, equation (5.6) yields the estimate

$$g^{1+1/n}(x) = g(x) + \mathrm{O}\left(\frac{1}{n}\ln|1-2x|\right) \tag{5.7}$$

which conforms to the intuitive expectation that short intervals of time are correlated with small changes of state. All of these properties imply that $g^t(x)$ is a two-way flow in 'time' that generalizes the discrete dynamics in (5.1).

The formal resemblance of these expressions to classical mechanics can be accentuated by introducing a new function

$$g^t(x) = z(t) \qquad 0 \leqslant z \leqslant \tfrac{1}{2} \tag{5.8}$$

in which the notation highlights the traditional role of 'time' as the independent variable and supresses the (dimensionally essential) dependence on the initial state x_0. It is then easy to check that (5.5) and (5.8) imply that

$$[\ln(\sqrt{2}^{\sqrt{2}})]^{-2}\frac{\mathrm{d}^2 z}{\mathrm{d}t^2} = (2z-1)\{[\ln(1-2z)]^2 + \ln(1-2z)\} \tag{5.9}$$

which re-expresses the dynamics in a form reminiscent of Newton's '$F = ma$'.

At first glance, it seems that equation (5.9) implies a time reversibility of the dynamics because the second derivative remains unchanged when t is replaced by $-t$. But, of course, the symmetries of differential equations are not necessarily shared by their solutions (i.e., 'spontaneous' symmetry breaking); and, in this case, it follows directly from (5.5) that when $0 < x < 1$,

$$\lim_{t\to\infty} g^t(x) \to \frac{1}{2} \qquad \text{whereas} \qquad \lim_{t\to-\infty} g^t(x) \to 0. \tag{5.10}$$

This asymmetric behaviour affects the identification of the mathematical backward flow, $g^t(x)$, $t < 0$, as a physical time reversal. Specifically, equation (5.6) implies that

$$g^{-1} \circ g = g^0 \tag{5.11a}$$

and also

$$g^0 \circ g^0 = g^0. \tag{5.11b}$$

However, despite the notational resemblance to f^{-1} in (5.3), g^{-1} does not signify an inverse function because the non-monotonic logistic map (5.4) has no inverse. This is correlated with the technical point that any function satisfying (5.11b) is said to be idempotent and is not necessarily equal to the identity function j. In the present example (5.5) shows that g^0 is actually a restriction of j. The general conclusion is that dynamical systems satisfying (5.11a) and (5.11b), where $g^0 \neq j$, do not correspond to time reversal. Quite the contrary, in these cases (5.11a) describes the *non-restorability* of states.

5.3. Index time: labelled sequences of configurations

After dispensing with the tightly constrained machinery of functional iterations in (5.1), all that remains is the weaker notion of aggregates of states that may be ordered by means of some identifying tag. In a stimulating article 'The Emergence of Time and Its Arrow from Timelessness', J B Barbour argued that this kind of 'relative configuration space of the universe' is actually sufficient to provide a complete description of events (pp 405–14 of [47]). He introduced this drastic demotion of time by quoting the results of a straw poll conducted among 42 participants at the Conference whose Proceedings appear in [47] (Magazon, Spain, 1991). The question was:

> Do you believe time is a truly basic concept that must appear in the foundations of any theory of the world, or is it an effective concept that can be derived from more primitive notions in the same way that a notion of temperature can be recovered in statistical mechanics?

Twenty responded with the opinion that there was no time at all at a fundamental level; ten believed that time did exist at some basic level; and the rest abstained. Obviously, the largest group among these experts shared Poincaré's doubts concerning the fundamental significance of time.

If time is indeed a redundant concept then, contrary to the traditional view that different configurations are realized *at* different instants of time, it is Barbour's basic contention that

they *are* the instants of time. As a corollary, intervals of time are then derived as measures of the differences between configurations—which may be normalized by reference to a clock. Parallel processing computers form a class of practical devices whose operations precisely model Barbour's approach [61]. An obvious means of synchronizing multiple processors running simultaneously is to require that the entire system be controlled by a master clock so that a definite time can be ascribed to every elementary switching operation regardless of which processor is involved. However, since this method is basically equivalent to linearizing the time-stream of computations, it cannot achieve the speedup theoretically possible with parallel processing. The programming schemes used in practice avoid these wasteful references to a common (artificial) time base by organizing the computations as logic streams whose interconnections depend only on their relative order. In this setting, Barbour's relative configurations correspond to 'key points'—or, sets of signals within circuits—which are activated by 'cones of logic' originating from prior computations [62].

There are, however, many experimental situations involving quantum mechanics where it is not possible to distinguish between contiguous configurations. A prime example that updates Schrödinger's concerns regarding the operational meaning of quantum time is photon tunnelling through micron-thick dielectric mirrors [63–65]. The key result inferred from two-photon interference measurements is that the apparent tunnelling velocity of single photons traversing a dielectric barrier may be superluminal, $(1.7 \pm 0.2)c$. As might be expected, this observation has reignited long standing controversies among partisans of different measures of 'microtime' such as Larmor time, phase time, group-delay time, etc; and has also exposed ambiguities in the interpretation of experimental procedures [65, 66]. In any event, it is clear that quantum mechanics has had a contradictory influence: on the one hand, atomic clocks are the best means for obtaining ever more refined 'configuration pictures' that are the raw material for the Barbour program; but on the other hand, indecomposable processes, such as barrier tunnelling, show that Zeno's arrows still carry a sting.

And, speaking of arrows, where is the arrow of time in this amorphous scenario? Barbour's conjecture

> ... is that the ultimate origin of the arrow of time is the asymmetric structure of the configuration space of the world. ... Potentials are always represented as hills, valleys, wells, and walls. The configuration space of the world must be criss-crossed by the most extraordinary mountain ranges, ocean troughs, odd-shaped obstacles, and so forth. Above all, superimposed on everything is one prevailing direction, arrow if you like ... (p 411 of [47]).

No doubt this is an eloquent and sweeping proposal intended to provide '... a genuine explanation of the arrow of time ...'. But lacking any substantive evidence, it remains an aspiration rather than a blueprint for specific research. Nevertheless, this train of thought may be continued along more promising lines by switching to a qualitatively different kind of time.

5.4. Process time: time's arrows as emergent properties of complex processes

The generous distribution of arrows in figures 1(*b*), 2 and 3 shows that even simple gradient systems can exhibit a great variety of irreversible transitions. These diagrams also serve as reminders that irreversibility is not an elementary property of a single state or a grouping of states, but rather characterizes the processes that transform states into other states. The appearance of so many 'arrows of time' in these figures is due to the fact that these diagrams are not merely energy landscapes, but also incorporate—in the simplest possible way—the flow of transformations induced by external agencies such as hysteresis variables [67]. In other words,

the asymmetries conjectured by Barbour are evident in these diagrams simply because these representations contain more information than the configuration spaces considered previously in section 5.3.

In complex systems, it is generally not possible to associate a unique arrow of time with a set of transformations. For instance, if a state A is irreversibly transformed to a state B, and B is changed back again to A by other means that ensure that the restoration is perfect, then evidently the reset state A has no memory of the excursion to B, and the overall process does not generate an intrinsic arrow of time—even though technical time will keep on ticking throughout the interchanges $A \to B \to A$. This situation is illustrated by the deformation of an elastic–plastic torsion spring discussed in section 4.2. In this case, A is represented by the initial state, $\Theta = 0$, in equation (4.3), and B corresponds to the state of the system after plastic yield. The basic point of this example was to demonstrate that the irreversible $A \to B$ transformation was inherently simpler than the $B \to A$ restoration. Suppose, now, that a set of torsion springs were equally apportioned among states A and B. Since this spread removes the prejudicial effects of initial conditions, the only asymmetry available to influence any ensuing dynamics is that the $A \to B$ transitions are simpler to manage than $B \to A$. Under these circumstances it is plausible to assume that an arrow of time points in the 'easy' direction $A \to B$. An asymmetry principle of this kind is roughly analogous to the preferential population of states favoured by statistical weights or the greater sizes of capture basins on energy landscapes; but complexity measures suited for ranking trends among sets of processes are qualitatively more complicated [68].

The entire series of examples outlined in sections 1–4 illustrates the parallel development of complexity and the associated emergence of arrows of time. The initial example—a Hamiltonian invariant under time reversal (1.1)—represents a null case: here there are no arrows of time because all possible dynamics have time reversed complements of equivalent complexity. But, as indicated previously, not all null cases are tautologies. Hibernation, dormancy, and suspended animation all correspond to a partial or complete blunting of time's arrows in biological systems. Apart from the general slowing of chemical reactions due to decreasing Arrhenius factors at low temperatures, these suppressions of time are linked with complexity and technology. For instance, the successful reanimation of single frozen cells goes back 50 years [69], whereas assisted reproduction with cryopreserved blastocysts ($\gtrsim$200 cells) became available only within the last decade [70]. Obviously, both technical time as well as process time depend on the state of the art.

Arrows of time can also be switched on and off in quantum system. Measurements of 'old' muonium (μ^+e^-) [71, 72], time-selected atomic double resonance [73], and time-dependent Mössbauer spectra [74, 75] all confirm that resonances associated with excited states of quantum systems become narrower with increasing age. This trend is, of course, consistent with the time–energy uncertainty relations, but the underlying connections are not trivial since refined analyses show that there are no unique correlations between the time distributions of decays and energy spectra [76, 77]. The progressive narrowing of resonances suggests that the duration of the excited states is associated with an intrinsic time scale or arrow. In contrast, quantum systems can persist indefinitely in their ground states without any signs of ageing or arrows of time. By means of suitable population transfers between ground and excited states, it is possible to turn the arrows of time on and off, and thereby gain some insight into the physical meaning of 'time in' versus 'time out' in quantum systems. A clean test case is provided by the double-resonance fluorescence of a single trapped ion in an arrangement where the strong fluorescence can be used to monitor transitions of the weak fluorescence—this is the essence of Dehmelt's shelved electron scheme [78]. Suppose that initially an electron is promoted from a ground state to a long-lived excited state, say, the $^2D_{5/2}$ level of $^{198}Hg^+$ with a half-life of 0.1 s

[79, 80]. If after a time interval $\Delta t_1 \lesssim 0.1$s no decay has occurred, transfer the electron back to the ground state with coherent light pulses [81, 82], and wait for another time interval Δt_2. Then transfer the electron back up to $^2D_{5/2}$ with another set of π-pulses. Suppose that this second shelving lasts for a time Δt_3 until spontaneous emission finally returns the electron to the ground state $^2S_{1/2}$. In analogy with the preceding, the theoretical expectation is that long shelving times will be correlated with a subset of extremely sharp spectral lines [80, 83, 84]. Similar effects are routinely exploited in the construction of some atomic clocks. However, in the present case, the effective duration of the shelving time has been deliberately complicated by the adiabatic population transfers between $^2D_{5/2}$ and $^2S_{1/2}$. Although it seems obvious that the length of the intermediate time out (Δt_2) in the ground state should have no effect on the spectral narrowing, the junction between Δt_2 and Δt_3, i.e. the transition between time out and time in, is not clearly defined. Furthermore, if Δt_2 is sufficiently short, a coherent memory of Δt_1 might survive to influence the final decay spectrum.

Acknowledgments

Most of the results concerning the irreversible behaviour of magnetic cooperative systems are excerpted from research carried out in collaboration with B N Harmon, H G Latal, and R Olenick. We are grateful to S A Guralnick and A Sklar for many stimulating and informative discussions. TE also thanks the Enrico Fermi Institute of the University of Chicago for its hospitality.

References

[1] von Helmholtz H 1887 *J. Reine Angew. Math. (Crelle)* **100** 137–66
von Helmholtz H 1887 *J. Reine Angew. Math. (Crelle)* **100** 213–22
[2] Poincaré H 1889 *C. R. Acad. Sci., Paris* **108** 550–3
[3] Onsager L 1931 *Phys. Rev.* **37** 405–26
[4] Lamb S W J and Roberts J A G 1998 *Physica* D **112** 1–39
[5] Carathéodory C 1919 *Sitzungber Preus. Akad. Wiss. Math.-Phys. Kl.* 580–4
[6] Misra B 1978 *Proc. Natl Acad. Sci., USA* **75** 1627–31
[7] Olsen E T 1993 *Found. Phys. Lett.* **6** 327–37
[8] Boltzmann L 1895 *Nature* **51** 413–15
[9] Krylov N S 1979 *Works on the Foundations of Statistical Physics* (Princeton, NJ: Princeton University Press)
[10] Khinchin A I 1979 *Mathematical Foundations of Statistical Mechanics* (New York: Dover)
[11] Borel E 1912 *Book of the Opening of the Rice Institute* vol 2 (Houston, TX) pp 347–77
[12] Perrin J 1923 *Atoms* (New York: Van Nostrand)
[13] Denjoy A 1975 *Arnaud Denjoy: Evocation de l'Homme et de l'Oeuvre* (Paris: Société Mathématique de France) pp 21–3
[14] Mandelbrot B B 1983 *The Fractal Geometry of Nature* (New York: Freeman)
[15] Bernstein B, Karamolengos M and Erber T 1993 *Chaos Solitons Fractals* **3** 269–77
[16] Ewing J A 1900 *Magnetic Induction in Iron and Other Metals* (London: Electrician)
[17] Thomson J J 1904 *Phil. Mag.* **7** 237–65
[18] Gallop E G 1897 *Messenger Math.* **27** 6–12
[19] Erber T, Latal H G and Harmon B N 1971 *Advances in Chemical Physics* vol 20, ed I Prigogine and S A Rice (New York: Wiley) pp 71–134
[20] Weinstock H, Erber T and Nisenoff M 1985 *Phys. Rev.* B **31** 1535–53
[20a] Marsden J E and McCracken M 1976 *The Hopf Bifurcation and its Applications* (New York: Springer)
[20b] Thom R 1966 *Stabilité Structurelle et Morphogènése* (Paris: Édiscience)
[20c] Poston T and Stewart I 1978, *Catastrophe Theory and its Applications* (London: Pitman)
[20d] Gilmore R 1981 *Catastrophe Theory for Scientists and Engineers* (New York: Wiley)
[21] Vollmer J and Breymann W 1993 *Phys. Rev.* B **47** 11 767–73
[22] Erber T, Latal H G and Olenick R P 1981 *J. Appl. Phys.* **52** 1944–6

[23] Painlevé P 1904 *C. R. Acad. Sci., Paris* **138** 1555–7
[24] Wintner A 1941 *The Analytical Foundations of Celestial Mechanics* (Princeton, NJ: Princeton University Press)
[25] Seifert H and Threlfall W 1948 *Variationsrechnung im Grossen* (New York: Chelsea)
[26] Derrida B and Flyvbjerg H 1987 *J. Physique* **48** 971–8
[27] Flajolet P and Odlyzko A M 1990 *Eurocrypt 89 (Springer Lecture Notes in Computer Sciences vol 434)* ed J J Quisquater (Berlin: Springer) pp 329–55
[28] Edelman A and Kostlan E 1995 *Bull. Am. Math. Soc.* **32** 1–37
[29] Graham R L, Rothschild B L and Spencer J H 1990 *Ramsey Theory* (New York: Wiley)
[30] Erdös P and Rényi A 1960 *Magy. Tudom. Akad. Mat. Kutato Intézet Közleményei* **5** 17–61
[31] Bergson H 1922 *Durée et Simultanéité* (Paris: Alcan)
Bergson H 1965 *Duration and Simultaneity* (Indianapolis, IN: Bobbs-Merrill)
[32] Reichenbach H 1956 *The Direction of Time* (Berkeley, CA: University of California Press)
[33] Gold T (ed) 1967 *The Nature of Time* (Ithaca, NY: Cornell University Press)
[34] Grünbaum A 1973 *Philosophical Problems of Space and Time* (Boston, MA: Reidel)
[35] Davies P C W 1976 *The Physics of Time Asymmetry* (Berkeley, CA: University of California Press)
[35a] Smale S 1980 *The Mathematics of Time* (New York: Springer)
[36] Denbigh K G 1981 *Three Concepts of Time* (Berlin: Springer)
[37] Landsberg P T (ed) 1982 *The Enigma of Time* (Bristol: Hilger)
[38] Bennet C H 1982 *Int. J. Theor. Phys.* **21** 905–40
[39] Prigogine I and Stengers I 1984 *Order Out of Chaos* (New York: Bantam)
[40] Horwich P 1987 *Asymmetries in Time* (Cambridge, MA: MIT)
[41] Sachs R G 1987 *The Physics of Time Reversal* (Chicago, IL: University of Chicago Press)
[42] Zeh H D 1989 *The Physical Basis of the Direction of Time* (Berlin: Springer)
[43] Ebeling W and Engel H 1990 *Selbstorganization in der Zeit: Wissenschaftliche Taschenbücher (Math/Phys #309)* (Berlin: Akademie)
[44] Leff H and Rex A (ed) 1990 *Maxwell's Demon: Entropy, Information, Computing* (Princeton, NJ: Princeton University Press)
[45] Amis M 1992 *Time's Arrow* (New York: Harmony Books)
[46] Mackey M 1992 *Time's Arrow: The Origins of Thermodynamic Behaviour* (New York: Springer)
[47] Halliwell J J, Pérez-Mercader J and Zurek W H (ed) 1994 *Physical Origins of Time Asymmetry* (Cambridge: Cambridge University Press)
[48] Atmanspacher H and Ruhnau (ed) 1997 *Time, Temporality, Now* (New York: Springer)
[49] Savitt S F (ed) 1997 *Time's Arrows Today* (Cambridge: Cambridge University Press)
[50] Schulman L S 1997 *Time's Arrows and Quantum Measurement* (New York: Cambridge University Press)
[50b] Magnon A 1997 *Arrow of Time and Reality: In Search of Conciliation* (River Edge, CO: World Scientific)
[51] Poincaré H 1902 *La Science et l'Hypothèse* (Paris: Bibliothèque de Philosophie Scientifique)
[52] Poincaré H 1929 *The Foundations of Science* (New York: Science)
[53] Einstein A 1918 *Ann. Phys.* **55** 241–4
[54] Kretschmann E 1917 *Ann. Phys.* **53** 575–614
[55] Allan D W, Ashby N and Hodge C 1996 *The Science of Timekeeping* (Santa Clara: Hewlett-Packard Application Note 1289)
[56] Parkinson B W and Spilker J J Jr (ed) 1996 *Global Positioning System: Theory and Applications* vols 1 and 2 (Washington, DC: American Institute of Aeronautics and Astronautics)
[57] Schrödinger E 1931 *Sitzungber. Preuss. Akad. Wiss. Phys.-Math. Kl.* **12** 238–47
[58] Major F G 1998 *The Quantum Beat: The Physical Principles of Atomic Clocks* (New York: Springer)
[59] Anosov D V 1967 *Proc. Steklov Inst. Math.* **90**
[60] Sklar A 1984 Private communication
[61] Hockney G M 1993 Private communication
[62] Blackett R K 1996 *IEEE Spectrum* **33** 68–71
[63] Steinberg A M, Kwiat P G and Chiao R Y 1993 *Phys. Rev. Lett.* **71** 708–11
[64] Steinberg A M 1995 *Phys. Rev. Lett.* **74** 2405–9
[65] Steinberg A M 1995 *Phys. Rev.* A **52** 32–42
[66] Rényi A 1970 *Foundations of Probability* (San Francisco, CA: Holden Day)
[67] Erber T and Guralnick S A 1988 *Ann. Phys., NY* **181** 25–53
[68] Wackerbauer R, Witt A, Atmanspacher H, Kurths J and Scheingraber H 1994 *Chaos Solitons Fractals* **4** 133–73
[69] Karp G 1984 *Cell Biology* (New York: McGraw-Hill)
[70] Cohen J, Simons R F and Edwards R G 1985 *J. In Vitro Fertl. Embryo Transfer* **2** 59–64
[71] Casperson D E *et al* 1975 *Phys. Lett.* B **59** 397–400

[72] Boshier M G *et al* 1995 *Phys. Rev.* A **52** 1948–53
[73] Ma I-J, Mertens J, zu Putlitz G and Schütte G 1968 *Z. Phys.* **208** 352–63
[74] Wu C S, Lee Y K, Benczer-Koller N and Simms P 1960 *Phys. Rev. Lett.* **5** 432–5
[75] Neuwirth W 1966 *Z. Phys.* **197** 473–504
[76] Gromes D and Petzold J 1964 *Ann. Phys.* **14** 353–83
[77] Kerler W and Petzold J 1965 *Z. Phys.* **186** 168–89
[78] Dehmelt H G 1975 *Bull. Am. Phys. Soc.* **20** 60
[79] Bergquist J C, Hulet R G, Itano W M and Wineland D J 1986 *Phys. Rev. Lett.* **57** 1699–702
[80] Erber T, Hammerling P, Hockney G, Porrati M and Putterman S 1989 *Ann. Phys., NY* **190** 254–309
[81] Laine T A and Stenholm S 1996 *Phys. Rev.* A **53** 2501–12
[82] Bergmann K, Theuer H and Shore B W 1998 *Rev. Mod. Phys.* **70** 1003–25
[83] Hegerfeldt G C and Plenio M B 1995 *Phys. Rev.* A **52** 3333–43
[84] Hegerfeldt G C and Plenio M B 1996 *Phys. Rev.* A **53** 1164–78