23. Before Langmuir's work brought the experimental value closer to Bohr's, the latter was thought to be ten times the actual value. See U. Hoyer, "Introduction" in N. Bohr, *CW* (1981) 132. For the other comparisons of theory and experiment, see H & K, 288–89.
24. See B. Brecht, "The Street Scene" in J. Willett, ed., *Brecht on Theatre* (London: Methuen, 1964).
25. Pace van Fraassen (1980), ch. 1. I may justly be criticized, in fact, for harping on just two criteria for success. In particular in emphasizing what can be called synchronic virtues, I have ignored those diachronic virtues which were the concern of two previous commentators on the Bohr atom, Ernan McMullin (1965) and Imre Lakatos (1980). I think that these diachronic virtues are parasitic on the synchronic ones, but that would need to be argued.
26. To take an obvious example, behind the same software can lurk many different kinds of hardware.
27. This argument was suggested to me by the closing argument in chapter 1 of van Fraassen (1980).
28. Thus Suppe: "A theory has as its intended scope a natural kind class of phenomenal systems. . . . In propounding a theory, one commits oneself to the existence of the phenomenal systems within the theory's scope." (1989, 97)
29. In fact hydrodynamical models of the atom were produced in the nineteenth century at a time when a lot of the smart money was on non-atomic theories of matter.
30. Earlier versions of this paper were presented in a seminar at the University of South Carolina, and read as a paper at the College of Charleston, S.C. I would like to thank those present on those occasions for many valuable comments and criticisms. I have naturally ignored the latter.

PHILOSOPHICAL TOPICS
VOL. 18 NO. 2, FALL 1990

Models, Chaos, and Goodness of Fit

Stephen H. Kellert, Mark A. Stone, and Arthur Fine
Northwestern University

I

INTRODUCTION

Science builds models to manage and to understand the phenomenal world. In choosing between competing models, a common standard is based on the idea of goodness of fit: that model is best which best fits the data. This paper looks at the idea of goodness of fit for models of scattered data. We concentrate the discussion on a context of inquiry that is of considerable philosophical interest and explore the question of whether goodness of fit is decidable there.

The context for our discussion is chaos theory, where we focus on Robert Shaw's illuminating analysis of the dripping faucet as a chaotic system (1984). Shaw's analyis is comprehensive. It proposes not just a chaotic model of the dripping faucet, but a whole methodology for modeling complex and apparently aperiodic behavior. It raises the question of the suitability of chaos theory as a scientific approach. In examining this question, we argue for several theses: first, that with important exceptions, it is not generally decidable when a model constitutes a best fit; second, that goodness of fit in the contexts investigated by chaos theory has less to do with exact replication of the data than is typically assumed; finally, that the choice of whether to accept models like Shaw's depends on contextual and pragmatic judgments about methodological issues as much as on the goodness of fit of the particular model.

II

TWO TYPES OF MODELS

During the past two decades, under the name of "chaos theory," there has been an explosion of interest in the use of simple mathematical models to study complex behavior in physical systems, systems as diverse as hydrodynamic flow, autocatalytic chemical reactions, and population biology (Hao 1984; Holden 1986). Moreover, this "chaotic" approach promises fruitful applications to fields such as cardiology, economics, and neuroscience. One of the systems that provides an exemplary prototype for this approach is the familiar case of the dripping faucet. The analysis of this case offers some evidence that chaos theory can deliver on its promise.

As we all know, when water leaks slowly from the end of a pipe, it forms drops that eventually detach and fall, and at intervals that are (unfortunately!) detectable. For low rates of water flow, this dripping can be as regular as a metronome, exhibiting periodic behavior. As the flow rate is increased, however, the drops begin to fall irregularly. The transition from regular drips to irregular and seemingly patternless behavior suggested to Robert Shaw that study of the dripping faucet might provide insight into a similar phenomenon, the problem of the onset of turbulence in fluid flow past an obstacle.

Turbulence is along standing problem for classical physics. There is still no adequate classical treatment of the whorls and eddies that appear in waterfalls, whirlpools, and wakes. Before chaos theory, the standard approach was the account suggested by Lev Landau (1944). The Landau model seeks to understand turbulence by describing how smooth laminar flow (i.e., flow where the fluid can be considered as moving along in separate, neatly stacked sheets) becomes disrupted as the speed of the fluid past an object is increased. To illustrate the technique, imagine water in a slowly moving creek flowing past a large rock. Suppose we place a very accurate device downstream from the rock that measures the velocity of the water at that point. For laminar flow, the device will register a constant value, but as the current increases in speed the smooth flow lines around the rock begin to bend,

causing undulations which detach into small eddies that move downstream in time. As one of these eddies passes our measuring device, the velocity will register an increase, then a decrease, and then it will return to the undisturbed value until another eddy comes by. The sequence of velocity measurements—the time series—changes from constant to periodic behavior (see Fig. 1a,b).

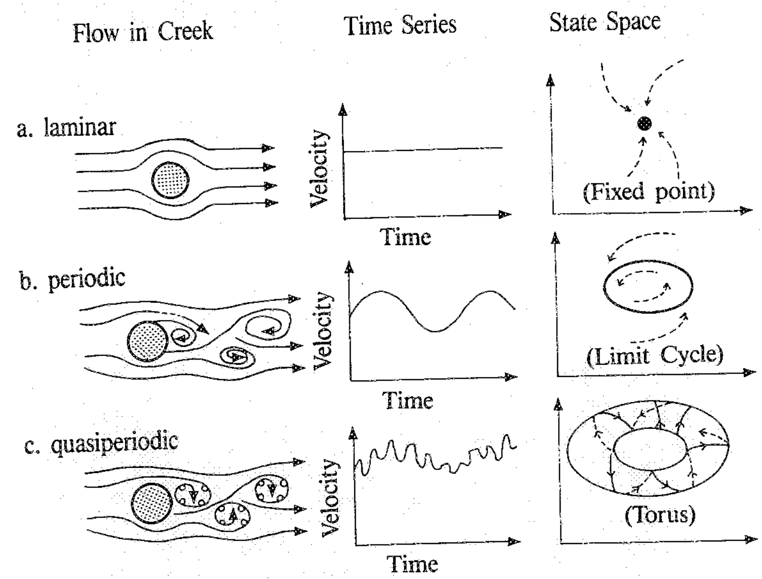

FIGURE 1. (after Kadanoff (1983))

So far, a straightforward model suggests itself for this system. At low flow velocities, the rock has no susbstantial effect; small perturbations in velocity are quickly smoothed out. We can plot this in state space; that is, on a graph where a point represents the velocity-state of the system at a given moment. The initial flow velocity then represents what is called a *fixed point attractor*, which is just to say that the flow velocity remains constant and, if perturbed, quickly returns to the initial value. At higher velocities, when the rock causes eddies to appear below it, we get periodic flow characterized by a *limit cycle attractor*. That is, in state space, we have a circle representing the range of velocities through which the state of the system cycles. As eddies appear within eddies, however, the time series of velocity measurements will vary with two frequencies, as seen in Figure 1c. The point representing the state of the system in state space now spirals around a two-dimensional torus. If the two frequencies are not rationally commensurate, the system will never return exactly to the same state; instead it will wind around and around the torus forever and never regain its starting point. This situation is labeled *quasi-periodic behavior*.

The heart of the Landau model then is the torus. All systems of fluid flow, even turbulent systems, are modeled by an n-dimensional torus (where the dimensionality n depends on the system of eddies within eddies). One treats behavior that appears to be aperiodic as quasi-periodic, making some or all of the frequencies incommensurable. Since it is not obvious how to distinguish genuinely random turbulence from highly complex quasi-periodic behavior, rather than treating a particular fluid system as behaving randomly, it seems that we could always look for a Landau model with suitable frequencies and dimensionality. The Landau model for the onset of turbulence thus embodies a classical view of complexity, which treats complex behavior as arising from a welter of competing influences, degrees of freedom, or interacting subsystems. (Below we speak of a "Landau model" to refer, generically, to any model that adopts this approach.)

Chaos theory challenges this classical picture of complexity and randomness.[1] One of the seminal works in chaos theory is an alternative account for the onset of turbulence. Known as the Ruelle-Takens-Newhouse (RTN) model, this account questions the idea that the complex behavior in turbulent flow must be modeled by the agglomeration of incommensurable frequencies (Newhouse, Ruelle, and Takens 1978). Instead, the RTN model explains the transition to turbulence by the appearance in state space of an attractor that represents extremely complicated dynamical behavior, yet is described by a very simple set of mathematical equations. This novel mathematical object is called a "strange attractor."

In the RTN model, the behavior of fluid flow past an obstacle follows the path laid down by Landau only up to the appearance of a two-dimensional torus. After that point, a further increase in the flow rate can render this attractor unstable, and for a wide variety of cases, the behavior will change to weak turbulence characterized by motion on a strange attractor. The strange attractor is characterized by several important features: (1) it is an attractor, that is, an object with no volume in state space and toward which all nearby points converge; (2) typically, it has the appearance of a fractal, which is to say, a stack of two-dimensional sheets displaying a self-similar packing structure; (3) motion on it exhibits sensitive dependence on initial conditions, that is, for any point on the attractor there is a nearby point which traces a path that diverges exponentially from that of the first; (4) it can be generated by the numerical integration of a very simple set of exact dynamical equations.

We now turn to the dripping faucet for some lessons in how to adjudicate the choice between the torus and the strange attractor.

III

THE DRIPPING FAUCET

Shaw's analysis of the dripping faucet is an exemplar for chaos theory. It describes the experimental procedures for obtaining relevant data from a physical system and develops the theoretical apparatus needed to analyze that data using the mathematics of chaos. The analysis is especially important in two regards. First, it shows how to study a very complex physical process using simple chaotic mathematical models. This vindicates the notion that exotic mathematical objects, like strange attractors and fractal sets, are relevant to physical systems in the natural world. Second, the analysis develops an invariant measure for the unpredictability of chaotic systems: the "entropy measure." This measure provides a quantitative tool for comparing different chaotic systems and for characterizing the way they respond to changes in their environment.

In the analysis of the dripping faucet system, the crucial first step is to focus attention on just one macroscopic feature of a complicated situation. Instead of trying to analyze the changing shape of the drops as they form and detach, data are collected and presented as a series of numbers (T_1, T_2, T_3 . . .) that represent the elapsed time between successive drops (the "drop interval"). The problem of the transition to complex behavior then takes the form of modeling the change from a situation where all these numbers are the same to a situation where they vary in a complex and seemingly random fashion.

By plotting the pairs (T_n, T_{n+1}) Shaw and his colleagues could probe the data collected for evidence that a simple mathematical model would be appropriate for understanding the behavior. In the case of periodic dripping, T_n is always (roughly) equal to T_{n+1} and the graph of the data is approximately a point (see Fig. 2a). After the flow rate is increased into the aperiodic region, we would expect to see a plot such as that in Figure 2b, if we assume the behavior to be random. That is, if there were no correlation at all between successive drop intervals, the graph would show a random scatter of points. Such a scatter could be the result of some intrinsically probabilistic mechanism at work, or it could point to the presence of competing influences from innumerable microscopic subsystems of the Landau type (the nested "eddies").

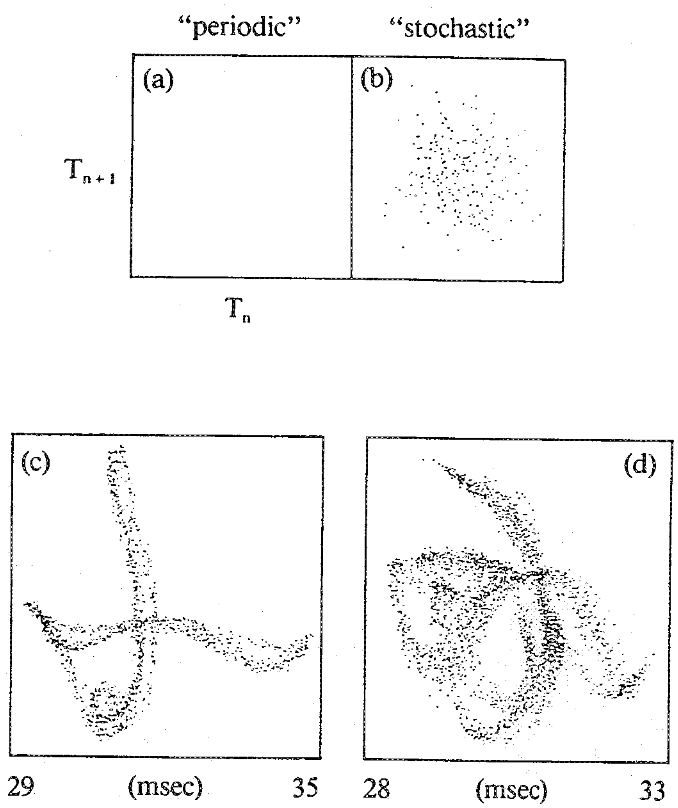

FIGURE 2

Surprisingly, the experimental data produce plots such as those in Figures 2c and 2d. These patterns found in data from "random" dripping help explain why researchers often speak of finding "order in chaos." This simple graphical analysis suggests that a complicated continuous system, which presumably would require a high dimensional Landau model, can "behave in a 'low-dimensional' fashion" (6)[2].

Following this suggestion, the model Shaw proposes is quite different from a Landau-type one. It is based on a simple analogy, which constitutes the second stage in the analysis. Suppose we have a body suspended from a spring, and that the body undergoes a constant increase in mass. As the mass increases, the body will accelerate downwards. Let us suppose that at some critical distance the body separates into two bodies—one with the original mass still attached to the spring, which will then spring back up to its original position, and the other body with some mass Δm that detaches and falls (see Fig. 3). If g is

the acceleration due to gravity and k is the spring coefficient, then the system can be characterized as follows (15):

$$mg - kx \approx m\ddot{x}$$
$$\dot{m} = \text{constant}$$
$$\Delta m \propto \dot{x}|_{x=x_0}$$

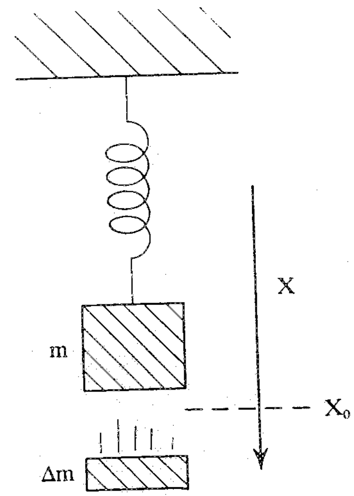

FIGURE 3

The assumption in the analogy is that the viscosity of water acts like the spring coefficent and that flow rate is analogous to the rate of mass change. Thus the model embodies the hypothesis that, with respect to the effects of viscosity and flow rate, dripping water behaves like a mass on a spring.

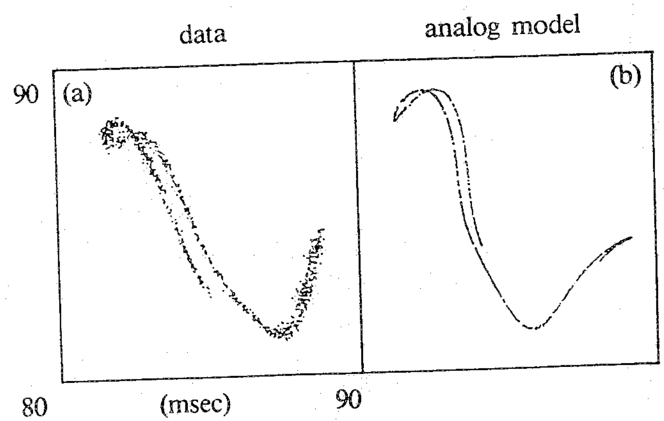

FIGURE 4

To examine the hypothesis, Shaw constructed an analog simulation of the mass and spring and used it to generate time series data, just as with the dripping faucet. The result was that "physical faucet data can be found which closely resembles time vs. time maps obtained from the analog simulation . . . "(16). This resemblance (see Fig. 4) seems to confirm the hypothesis. The advantage of postulating the dripping faucet as like a mass on a spring is that it "makes available continuous variables, which are not obtained from the physical experiment as it is presently conducted"(17). Pursuing the analogy, the functions describing the mass and spring are taken to indicate the underlying dynamic in the dripping faucet and used to construct an attractor in state space for the system.

The result, however, is not an n-dimensional torus. Instead we get a strange attractor (see Fig. 5). According to Shaw, "this structure can be recognized as *Rössler attractor* in its 'screw type' or 'funnel' parameter regime. The close correspondence of model and experiment . . . argues that such a structure is imbedded in the infinite-dimensional state space of the fluid system" (17). The dripping faucet experiment thus demonstrates how strange attractors may model real physical systems.

Specification of the attractor yields a geometric specification of the model. If the same attractor manifests itself in other systems, then we would have reason to believe that the same type of chaotic behavior is at work in these other cases. The fact that the dripping faucet data "looks" as though it matches the Rössler attractor hardly constitutes a sufficient test of the model, however. It would also be useful to have a quantified characterization of the dripping faucet model.

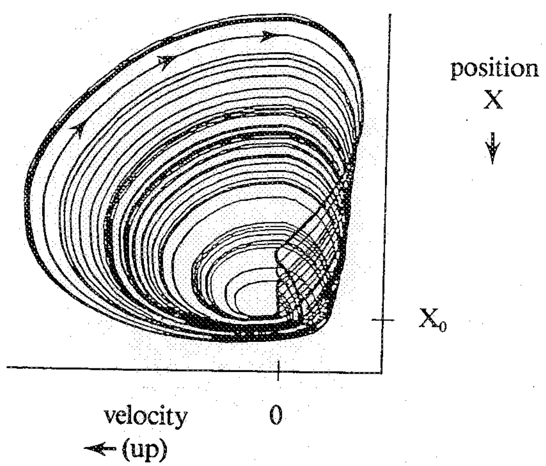

FIGURE 5

The third stage of Shaw's experimental analysis develops a method for characterizing strange attractors by giving a quantitative measure for their unpredictability. He contends that the resulting measure comes from data that contains a stochastic element beyond what would be introduced by observational error alone. Shaw borrows from information theory to develop this measure, which he calls the "entropy" or "rate of information production" of the system.

Laplace once conjectured that, given precise information about the present position and momentum of every particle in the universe, it would in principle be possible to predict exactly the state of the universe at any future time. However untenable Laplace's idea may be, it will serve as a useful vehicle in order to illustrate the idea of an entropy measure. Suppose that we have a three-part system of transmitter (T), channel of transmission (C), and receiver (R). On the one hand, we might consider the physical world to be an instance of this system, where present states (T) are transmitted via the laws of nature (C) into future states (R). On the other hand, we might consider a scientific theory to be an instance of this system, where representations of present states (T) are transmitted via a model (C) into predictions (R). Laplace's conjecture amounts to the claim that our theories could in principle be perfect transmitters, so that predictions match future states perfectly. If perfect predictability fails, according to Laplace, it can only be either because our model is imperfect or because our representation of present states is imperfect.

Both these ideas can be quantified. Given initial data, we can measure the degree of predictive accuracy of a model and compare degrees of accuracy of different models. But we can also measure the effect of increased accuracy in our initial data; given a model, we can measure how much our predictive accuracy is improved by improving accuracy in our initial data. Call the first of these the *measure of predictability* (P); it measures degrees of predictive accuracy. Call the second of these the *information measure* (I); it measures the impact of adding to our initial information. Their difference ($P-I$) represents the rate at which predictive accuracy is lost over time. This is the *entropy measure*.

Systems governed by classical attractors, like the fixed point and limit cycle attractors, do not have a positive entropy measure. Their predictive accuracy does not degenerate over time. Chaotic systems governed by strange attractors have a positive entropy measure; they lose accuracy over time at an increasing rate. Furthermore, the entropy measure is an invariant of the particular type of chaotic system. It is thus a signature of the system just as much as the attractor is, only a quantifiable signature rather than a geometric one. However, to get a model with positive entropy, one could take a classical system and simply add a noise term. Shaw contrasts such an alternative, the parabolic map, with his chaotic model.[3] With noise added, the parabolic map can produce a point spread that roughly matches the data and one for which the entropy measure is positive.

The issue is the source of noise; classically one looks to measurement error as the source. According to Shaw, however, the data from the dripping faucet indicate that not all of the noise in the system can be attributed to measurement error alone (79–83). So in particular, we cannot interpret the noise term in the parabolic map as representing just measurement error. We must look for another source. By contrast, Shaw's chaotic model is complete. The assumption that the system is chaotic suffices to account for all the noise. We can agree with Shaw that this fact seems a more conclusive reason for thinking that chaos provides a good model of the dripping faucet than simply the fact that a state space graph resembles the Rössler attractor. Nevertheless, chaotic systems are not predictable in any standard sense, and models of chaotic systems do not give us a comprehensive set of predictions that we can verify to confirm the model. The entropy measure is the only exact quantitative prediction that Shaw's technique generates. Thus we must ask two related questions: is the dripping faucet in fact a chaotic system and, more specifically, is Shaw's chaos model superior to a Landau model of aperiodic flow? Shaw argues that the latter question is fundamentally undecidable. We turn now to an assessment of Shaw's argument.

IV

MODELING AND PREDICTIVE OPTIMALITY

Shaw's argument presupposes an account of science as model-building. More specifically, this account of science sees physical systems as producing "data streams," or sequences of numbers, which the scientist seeks to match by constructing a model in the form of an algorithm that also generates a number sequence. One can then try to judge these models by how closely their data stream mimics the stream produced by the physical system, as well as by other criteria such as simplicity or explanatory power. This "two-stream" account agrees with a picture of scientific theorizing proposed by Ray Solomonoff in 1960 described as follows by Gregory Chaitin (1975, 49): "a theory that enables one to understand a series of observations is seen as a small computer program that reproduces the observations and makes predictions about possible future observations. The smaller the program, the more comprehensive the theory and the greater the degree of understanding."

The two-stream doctrine suggests a straightforward way to quantify goodness of fit in terms of what we will call *predictive optimality*. As Shaw describes it (88), we take a string of data from the actual system starting with some value X_0 and compare it to the string generated by the model when it is supplied with X_0 as the initial condition. The two strings are stipulated to be identical at time zero, but they will eventually begin to differ. The slower the rate at which this difference grows, the more predictively optimal the model is. As Shaw writes, ". . . the average number of matching digits, and the average rate of loss of matching digits as predictions farther into the future are attempted, can be recorded. It is unclear to the writer whether such measures are coordinate independent, but they seem operationally well-defined in a given setting, and capable of measuring *improving* of a model."

Predictive accuracy is not the only thing we value in a model, however. Other things being equal, a model should be *concise*. Good predictions are of less value if the computational apparatus required to generate them is extremely unwieldy. The conciseness of a model can also be quantified. The computational apparatus is an algorithm, and we can measure the computational complexity of the algorithm (Kolmogorov 1965). Obviously there is a tradeoff between predictive accuracy and computational simplicity. We can always obtain a simpler algorithm at the expense of accuracy. It might also seem that greater accuracy could always be obtained if we were willing to resort to more complex algorithms. But if the system under study contains some genuinely random elements, then predictive accuracy is limited and no algorithm can duplicate the output stream of the system exactly and reliably.

When, then, is a model optimal? The difficulty is deciding when we have reached the point of diminishing returns in sacrificing simplicity for accuracy. In the extreme, the difficulty is deciding when greater accuracy is not obtainable; in deciding, that is, when to model a system as though it contains genuinely random elements.

V

UNDECIDABILITY

The appearance of randomness in chaotic systems poses a strategic problem for theorizing. If a novel analysis of apparently random data can reveal intelligible patterns, how can we be justified in ceasing the search for simple models? At what point do we regard a statistical spread in the data as not requiring further analysis? This is equivalent to the question of predictive optimality. Consider the plot in Figure 4a once more. There is a specific squiggly "shape" to the data that is apparent to the eye, and it is indeed this geometric pattern that the model successfully matches. But there is also some "fuzziness," representing a statistical spread about the specific squiggle. To say we have successfully modeled the data can only mean that we have successfully modeled the data stream *so far as it is not random.* The remaining spread must then be regarded as mere noise.

At some point, one must stop looking for "fine structure" and start seeing "random noise." But when? For chaotic systems, Shaw argues that this question is undecidable. In principle, he contends, there is no way to determine which fluctuations are "signal" and which are "noise."

The argument proceeds from the fact, noted above, that a model of a chaotic system will necessarily produce a data stream that differs from the system.

> The discrepancy between the predictions of a model and observations of a physical system is, to the observer, a random element. If the model embodies the observer's complete knowledge of a system (including perhaps the addition of a noise element), he will be unable to distinguish the two resulting number streams, should he get them mixed up. A "proof" that the model was the best possible would involve a proof that the random element, or discrepancy, was intrinsic to the system, or "truly random," not susceptible to any further deterministic description. (93)

Shaw suggests that proving a number sequence to be random is an undecidable problem. More specifically, he means it is a search task not guaranteed to halt. One can look for an effective rule that generates the sequence and upon finding it know that the sequence is not random. Yet from the fact that one has *not* found such a rule, it does not follow that there is none. It does not follow that the sequence is in fact random.

"As Chaitin and others argue on the context of the theory of algorithms, one cannot 'prove' a number to be random, to possess such a proof would imply a logical contradiction. Thus, given a set of seemingly random numbers, there is no way in general to show that there does not exist an algorithm shorter than the list itself to generate the numbers, and thus reduce the randomness"(94). In discussing whether one can prove the assertion that a given model best fits the data, Shaw suggests (somewhat tentatively, to be sure) a connection

with Gödel incompleteness; i.e., the occurrence of undecidable sentences in any sufficiently rich system of formal arithmetic.

We shall try to bring out important ways in which the optimality of a model is a matter of non-algorithmic judgments that depend sensitively on specific features of the context. While in this sense it may be undecidable, we suggest that nothing so imposing as Gödel incompleteness lies behind it. Further, the undecidability is not so universal as Shaw's remarks would indicate. We turn to the latter point first.

VI

QUANTUM STATISTICS

Shaw rests his suggestion that the optimality of a model is undecidable on the claim that one cannot show that a system is truly random, in the sense of its being "not susceptible to any further deterministic description" (93–94). Investigations in the foundations of quantum theory, however, show that this claim is false. The most celebrated foundational result of the past two decades is Bell's theorem (Bell 1987), one version of which demonstrates precisely the impossibility of giving a deterministic model for a whole class of experiments with stochastic outputs. To show the connection to Shaw's problem, we describe the situation for just one experiment in the class. (See Cushing and McMullin for further references and discussion.)

In the experiment an atomic source steadily decays and emits pairs of spin-½ particles (e.g., electrons) in the singlet state (so, the total spin is 0, and it is conserved). After emission, the electrons in a pair separate and move in opposite directions to distinct wings of the experimental apparatus for observation. In the wings, there are instruments that can be set to measure the spin component of an electron along one of two directions in the plane perpendicular to the electron's line of motion. Thus the experiment involves the measurement of four variables, either A or A' (for one particle in a pair) and either B or B' (for the other particle in the same pair), corresponding to the four possible orientations for the measurement of spin. A particular spin measurement has two possible outcomes: the electron is either spinning clockwise ("up") or counterclockwise ("down") in the measured direction. We will record the "up" result with a "1" and the "down" with a "0". Under these conventions each of the four variables takes either 1 or 0 as a value. In four separate runs, on paired particles, we measure orientations A with B, then A with B', then A' with B, and finally A′ with B′. The best tested experimental geometry corresponds to the situation where the relative angle between directions in the first three runs (A-B, A-B′, and A′-B) is 135° and the relative angle between the orientations in the last run (A′-B′) is 45°. In a number of experiments of this type (more accurately, in experiments involving photon polarization, which is formally similar to electron spin), the statistics predicted by the quantum theory have been very well confirmed. Rounded to just one decimal place, the statistics are these.

$$\mathrm{P(A)=P(B)=P(A')=P(B')=0.5} \quad (1)$$

$$\mathrm{P(AB)=P(AB')=P(A'B)=0.4 \text{ and } P(A'B')=0.1} \quad (2)$$

(We write $P(A)$ for Prob($A=1$) and $P(AB)$ for Prob($A=1$ & $B=1$), and similarly for the other variables.) We shall see that, contra the claim made in connection with chaos theory, an experiment producing these statistics is "not susceptible to . . . deterministic description."

To show this, suppose (to the contrary) that there were factors that determined the experimental outcomes. Let "x" be a variable that ranges over these outcome-determining factors. If you like, one can think of x as an ordered couple whose two components constitute the factors relevant, respectively, to the two separate particles in an emitted pair. Thus if an emitted pair is characterized by a particular factor x and we measure A on one particle in the pair, then x (or its relevant component) determines whether the result of the A-measurement is either 1 or 0; similarly x would determine the outcome if we measure B (or B') on the other particle in the pair. With this understanding, to suppose that there is a deterministic description of the experimental outcomes amounts to representing the measurement outcomes by functions $A(x)$, $B(x)$, $A'(x)$, and $B'(x)$ taking values either 1 or 0, depending on the determining factor x. We will show that such a deterministic representation is inconsistent with the data in (1) and (2).

For that purpose, notice that if numbers q and r are either 0 or 1, then $qr \leq q$ and $qr \leq r$. If p is also either 0 or 1 then the following inequality holds

$$qr \leq pr - pq + q$$

since, depending on whether $p=0$ or $p=1$, the right hand side is either q or r. Because averaging over the values of variables preserves sums, differences and order, it follows that

$$\langle qr\rangle \leq \langle pr\rangle - \langle pq\rangle + \langle q\rangle$$

if p, q, and r are random variables taking only 0 or 1 as values, and $\langle . \rangle$ denotes the average (or expected) value of the enclosed variable.

If we now set $q=B(x)$, $r=B'(x)$ and $p=A'(x)$, then

$$\langle BB'\rangle \leq \langle A'B'\rangle - \langle A'B\rangle + \langle B\rangle.$$

If we set $q=A(x)$, $r=B'(x)$, and $p=B(x)$, then (after transposing)

$$\langle AB'\rangle + \langle AB\rangle - \langle A\rangle \leq \langle BB'\rangle.$$

Combining these two inequalities yields

$$\langle AB\rangle + \langle AB'\rangle + \langle A'B\rangle - \langle A'B'\rangle \leq \langle A\rangle + \langle B\rangle.$$

For variables that take only 0 and 1 as values, averages are the same as probabilities; i.e., $\langle AB\rangle = P(AB)$, and similarly for the other pairs. Thus we obtain the "Bell inequality":

$$P(AB)+P(AB')+P(A'B)-P(A'B') \leq P(A)+P(B).$$

From (2), the lefthand side equals 1.1, and from (1), the righthand side equals 1 in violation of the inequality. Thus the supposition that there are factors determining the observed values leads to the Bell inequality, which is inconsistent with the data in (1) and (2).

In addition to determinism, we should point out that the preceding demonstration involves an assumption about the measurement procedure. In representing the measurement outcomes as functions of certain determining factors x, we suppose that the outcome of the measurement of a variable in one wing, say A, is not affected by which of the variables (B or B') is measured in the other wing. If there were interference of this sort between the measurement performed in one wing and the outcome obtained in the other wing, we should have to represent the A-outcome at x not by $A(x)$ but by $A(x,B)$ or $A(x,B')$. In that case, however, the preceding argument would not go through. For similar reasons, we have to assume that the factor x (or its relevant part) that determines the outcome of a measurement occurring in one wing is not influenced by the measurement being carried out in the other wing. In the quantum mechanical experiment described above, we can arrange things so as to be reasonably certain that the spin measurements of the particles in an emitted pair are spacelike-separated; i.e., so that no subluminal influences between the two wings can create the unwanted interference. Were only a single system involved, however, there could be difficulties in preventing one measurement from interfering with the outcome of the other. (This is a problem with the arrangement in Home and Sengupta.)

These no-interference assumptions, however, do not blunt the force of the argument against undecidability. That argument is compelling provided there are physical systems producing data as in (1) and (2) in arrangements for which we can be reasonably sure that the no-interference assumptions are satisfied according to the best physics available. Against this background, the derivation of the Bell inequality shows that no deterministic account of the data is possible. Hence, contrary to the undecidability claim, the question of whether randomseeming processes can be modeled deterministically is sometimes decidable. The Bell theorem shows that its decidability is not so much a question of the general logic of randomness as one concerning the specific character of the experimental data. Thus the issue is more particular and less universal than one might have supposed.

It is easy to see, however, how one could be misled here. If we only attend to the sequence of outcomes in the measurement of one single variable, like the successive time intervals between drips in a leaky faucet then (in principle at least), however random-looking the sequence, one can always generate it as the output of a deterministic mechanism for we can always find a function from the integers to the output sequence and then suppose that the integers code for the outcome-determining factors. We can try to rule out such trivializations by adding requirements of computability and then bring to bear the technical construal of random sequences in terms of computational complexity (or perhaps some other account—e.g., that of time-complexity). Under such construals, as discussed in the preceding section, randomness may be undecidable for even a single sequence. Nevertheless, the Bell theorem shows that if we broaden our outlook to more complex experimental situations that involve several stochastic variables (four is the minimum number for the Bell theorem; see Fine) then, depending on the particular circumstances, deterministic modeling can entail constraints that the data simply fail to satisfy.

VII

THE PRAGMATICS OF CHOICE

What Shaw says about decidability is a polemic as much as an argument. Shaw argues that goodness of fit is undecidable because he *wants* it to be undecidable; he wants the methodology of chaos theory to be a live option to the traditional Landau model. If Shaw is a little overzealous in his invocation of Gödel undecidability, perhaps this is understandable. The name of Gödel carries weight.

Yet undecidability here is of humbler origins. At issue is the traditional curve plotter's problem with a stochastic twist: given a plot of data points, which of the infinitely many curves that can fit those data points is best? The twist for stochastic data is that the very notion of a data point may be ambiguous. Is a cluster of nearby data points to be treated as a single point scattered by the coarse-graining of our measurement and representation scheme? If so, then any curve that passes through the cluster is a good fit. Or is a cluster of nearby data points to be treated as a series of individual points all of which must fit the curve? In so far as there is no conclusive answer to these questions, there is no conclusive test for best curve.

Although all parties agree that models should be testable, part of the issue between Landau modeling and chaotic modeling concerns the very conception of a test. By traditional standards, the chaotic model of the dripping faucet is only weakly testable. Confirmation rests on the general fit of the Rössler attractor and the entropy measure. The traditional standard employed, however, equates testability with predictive optimality, conceived of as point by point matching. Shaw wants to loosen the grip of this traditional view. Thus Shaw suggests that for scattered data of the sort one gets from the dripping faucet, pattern matching rather than exact reproduction of the data points constitutes the right sort of test. Clearly if we *knew* we were dealing with inherently random elements, then pattern matching would be suitable, since there would then be no question of trying to achieve point by point matching of the two data streams. The polemic about true randomness being undecidable-in-principle seeks to reshape the conception of testability in terms of criteria appropriate for random processes.

The line of argument goes something like this. If we knew we were dealing with random elements, we would not insist on predictive optimality as our standard of what counts as a best fit. So long as we do not insist on matching predictions with the data on a point by point basis, however, chaotic models pass goodness of fit tests very well. They generate a close fit to the "shape" and "spread" of the data plot, as close a fit as can be produced by any Landau model. They also match the positive entropy measure exhibited by the data, a match that is beyond the reach of Landau models without special noise terms. Moreover, the argument continues, randomness is not decidable. For all we know, in the case of data like that obtained from the dripping faucet, we *are* dealing with random elements. Hence, it concludes, we should not insist on point by point matching, and should move instead to pattern matching and other more suitable criteria.

The last point reveals the flaw. For even in cases where we agree that one can not decide whether the data represent genuine randomness, it does not follow that we should assimilate our standards of testability to those of truly random processes. We could equally well go the other way! That is, undecidability leaves us with a choice of whether to treat the data set as though it contained random features, or to treat it deterministically. It leaves us with a choice of whether to insist on point by point matching as the basic criterion of goodness of fit for our models or to move to pattern matching and other criteria. Undecidability alone does not tilt the balance one way or the other.

So the polemic misfires. Nevertheless, Shaw's discussion of testability for models of the dripping faucet data points to something important; namely, the pragmatic equivalence between judgments of randomness and standards of goodness of fit. It suggests that there is no point in treating the modeling situation as though it involves two separate issues, an ontological one about "true randomness" and an epistemological one concerning goodness of fit. There is no point in supposing that we can separately decide about "true randomness" and then fix on appropriate standards of testability. From a practical point of view, the issues and the decisions are linked.

The quantum theory can teach us a lesson about these decisions. The history of the quantum theory records a series of premature congratulations over alleged "proofs" of the impossiblity of a more deterministic substratum out of which to concoct a rival theoretical approach (the various "no-hidden-variables" arguments). The Bell theorem finally moved the discussion beyond the level of self-congratulation by making it clear that the question of adequacy of a model needs to be relativized to a specific family of constraints for which there are reasonably clear formulations and experimental tests (the no-interference conditions discussed in section VI). In chaos theory, the question of "true randomness" is just the question of whether there are "hidden variables" in the faucet. The polemic about decidability shows that this question is tantamount to fixing on standards of goodness of fit. The lesson to learn from the quantum theory is that here too goodness of fit needs to be made relative to a specific family of relationships that can be judged reasonably well from the data. For modeling scattered data, the appropriate relationships are qualitative ones:[4] the patterns and shape of the plot, and associated measures like the spread and entropy measure. (A related move is contained in Dembski, who suggests that relativization to specifiable families of patterns might be a useful alternative to the current, intractable treatments of randomness itself.) In general, one would expect the relationships in the family to depend on the context of application, available background information, and on the uses one anticipates for the model. Clearly these involve judgments of practice as much as of principle, and this way of proceeding would make plain the extent to which the issue of goodness of fit relies on pragmatic judgments.

Other desiderata for choosing among competing models emphasize this pragmatic lesson as well. For example, Shaw's approach and the Landau approach will generate "types" of curves that will always differ in certain systematic ways. Broadly, the difference amounts to this: chaos theory gives us relatively simple models with a loose fit to

the data, while the Landau approach gives us more complex models that can fit the data more accurately, at least potentially. To be more specific, chaotic models have a positive entropy measure, while Landau models (without special noise terms) do not. This means that the rate of loss of accuracy of a particular Landau model will remain fairly constant over time, whereas chaotic models tend to lose predictive accuracy at an increasing rate. In principle, then, for any given chaotic model, there exists a more exact Landau model. However, Landau models of an interesting degree of accuracy have to be enormously complex. Thus although one may talk programmatically about the existence of more accurate Landau models, in practice one does not actually know how to find them. In these circumstances, the fact that chaos theory does the job of pattern matching with models that are much simpler than those of the Landau approach would seem to weigh in its favor, at least from a pragmatic point of view.

Another value for assessing models is scope. The Rössler attractor is a simple model of a very complex pattern of behavior. We can reasonably expect that the same attractor might be found in the data from systems of aperiodic flow besides the dripping faucet. Thus we expect the model to have broad scope, unifying various phenomena by assimilating them to the same pattern. Because of this, we feel that it has explanatory power. We would seem to have done good explanatory work if we could say of many diverse aperiodic systems that they all exhibit behavior like the Rössler attractor. By contrast, suppose that we could actually generate a Landau model of the dripping faucet. It is unlikely that a particular Landau model would be applicable to other aperiodic systems. Indeed it is unlikely that a Landau model of one dripping faucet would be applicable to another dripping faucet, since differences in size, shape, materials, viscosity, and surface tension would not amount just to revised initial conditions but also to a different set of fundamental equations. Lack of scope and unifying power may give us reason to question whether Landau models would do adequate explanatory work.

The general point is that goodness of fit has to work together with features such as simplicity, scope, and explanatory power in order to ground judgments of choice. These features are context dependent and pragmatic. We think this is as it should be.

VIII

COMPETING PARADIGMS

In some ways, the choice between Landau models and chaotic models is like the choice during the Renaissance between Ptolemy's and Copernicus' models of planetary motion. Ptolemaic methods gave rise to endless compoundings of deferents and epicycles upon orbits, much like the complex coupled periodic motions of the Landau models. Copernicus', and then Kepler's, models offered greater simplicity, if not necessarily greater accuracy (at least in terms of line-of-sight data). Although, in a sense, the Ptolemaic models could not be refuted (they could always be adjusted to save the phenomena) still that did not prevent their overthrow.

Shaw's argument for chaos in the dripping faucet challenges the picture of turbulence presented in the Landau model. It also challenges the very notion that the evaluation of scientific models proceeds by the simple comparison of two data streams. In choosing models, one cannot expect to sort the data into "signal" and "noise" by means of routine manipulations. The task of deciding how much of the data to model, and what aspects of it require modeling, is a task that calls for good scientific judgment. Carrying it out is not routine (or "decidable"). In particular there may always be innovations in the offing, such as chaos theory, that compel us to reconsider where we have drawn the line between signal and noise, between pattern and mere scatter. This includes reconsidering where the line is between chaos that reveals order and chaos about which there is nothing more to be said.

Today there are those who see chaos theory opening up a new scientific renaissance, promising to be as revolutionary as Copernicus' theory was in his time. However the issue is decided, *applicability* will feature as a central factor. In this regard, it is worth noting that the approach typified by the Landau model has been virtually barren of applications. It is also worth noting that the promised applications of chaos theory, to which we alluded earlier (see section II), must be regarded thus far as just that, promises yet to be fulfilled.

NOTES

Stephen Kellert's current address is Department of History and Philosophy of Science, Indiana University; Mark Stone would like to acknowledge support for this research provided by a postdoctoral fellowship from the History and Philosophy of Science Program at Northwestern University; and Arthur Fine's work on this paper was supported by NSF Grant DIR-8905571.

1. Though we will usually speak of randomness, it may sometimes seem that one could speak equally well of unpredictability. The two concepts are not to be equated, although the task of sorting them out is beyond the scope of this paper. (See Stone, 1989).
2. Page number references in the text are to Shaw 1984.
3. Specifically, Shaw considers the function $\dot{x} = rx(1 - x) + z$ for values of the parameter r that yield periodic variation in $\dot{x}$, and where z is the added noise term (55).

4. See Kellert for further discussion of how chaos theory can be said to provide scientific understanding of a qualitative nature.

REFERENCES

Bell, J. S. *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: Cambridge UP, 1987.

Chaitin, G. "Randomness and Mathematical Proof." *Scientific American*. May 1975: 47–52.

Cushing, J. and E. McMullin, eds., *Philosophical Consequences of Quantum Theory.* Notre Dame: U of Notre Dame P, 1989.

Dembski, W. A. "Randomness by Design." *Nous* (forthcoming).

Fine, A. "Correlations and Efficiency: Experimental Tests of the Bell Inequalities." *Foundations of Physics* 19 (1989): 453–78.

Hao, B. L., ed. *Chaos*. Singapore: World Scientific Publishing Company, 1984.

Holden, A. V., ed. *Chaos*. Princeton: Princeton UP, 1986.

Home, D. and S. Sengupta. "Bell's Inequality and Non-contextual Dispersion-free States." *Physics Letters* 102A (1984): 159–62.

Kadanoff, L. "Roads to Chaos." *Physics Today* December 1983: 46–53.

Kellert, S. H. "Philosophical Questions About Chaos Theory." Diss. Northwestern University, 1990.

Kolmogorov, A. N. "Three Approaches to the Quantitative Definition of Information." *Problemy Peredachi Informatsii* (in translation) 1 (1965): 3–11.

Landau, L. "On the Problem of Turbulence." *C. R. Academy of Science, URSS 44* (1944): 311–316.

Newhouse, S. E., D. Ruelle, and F. Takens. "Occurrence of Strange Axiom *A* Attractors near Quasi-Periodic Flows on T_m($m = 3$ or more)." *Communications in Mathematical Physics 64:* (1978) 35–43.

Shaw, R. *The Dripping Faucet as a Model Chaotic System*. Santa Cruz: Aerial Press, 1984.

Stone, M. A. "Chaos, Prediction and Laplacean Determinism." *American Philosophical Quarterly 26* (1989): 123–32.