Karl Popper (1961) is perhaps best known for reminding us that facts can falsify theories but cannot validate them. The theme of this paper derives from another, less well known, dictum of Popper's: Science does not aim at simplicity; it aims at parsimony. First, I will make some preliminary remarks about the goals of science. Then I will comment on the contributions of parsimony and simplicity to these goals. Finally, I will say something about the role that parsimony and simplicity have played in my own research, and in particular, my research in economics and statistics.

The Goals of Science

I will assume, for purposes of this discussion, that the scientist may be concerned with one or more of the following goals:

1. (Basic science) The central goals of basic science are (1) to describe the world (reality, if you please): both (a) specific facts about it ("The Earth revolves about the Sun each 365 1/4 days.") and (b) generalizations (laws) that describe collections of phenomena ("The periods of the planets around the Sun vary as the 3/2 power of their distances"
from it.

2. (Applied science) Laws connecting sets of variables allow inferences or predictions to be made from known values of some of the variables to unknown values of other variables. Inferences and predictions can be used, in turn, to invent and design artifacts (e.g., arches) that perform desired functions (support the weight and other stresses placed on them), or to anticipate and adapt to future events on the basis of knowledge about the present and past. At the times the predictions are to be realized, the new data can be used, of course, to test whether the laws continue to hold.

3. (Science as art) Although conforming to empirical truth is the central imperative of science, the working scientist, like every other professional, also responds to an esthetic imperative in which simplicity plays an important role. In general, beauty is thought (and felt) to lie in explaining much with little, and in finding pattern, especially simple pattern, in the midst of apparent complexity and disorder. Of course, these two aspects of beauty are closely connected: To be able to sum up a complex body of data in a relatively simple generalization (a pattern) is to explain much with little.

Daniel Berlyne (1960), in his research on curiosity, found that the length of time during which people will attend to a stimulus is a non-linear
function of its simplicity. A stimulus that is very simple (relative to the knowledge and experience of the viewer) will soon become boring: such obvious pattern as it possesses will soon exhaust its interest and attention will flag. If the stimulus is very complex (again, relative to the knowledge and experience of the viewer), no pattern will be recognized in it and attention will also soon flag.

Only if the viewer can continue, over some time, to detect new pattern in the stimulus, or to elaborate the pattern that has been detected, will his or her attention continue to focus on it. What is moderately complex relative to the viewer is interesting and beautiful because it can be perceived in terms of parsimonious patterns that can be extracted with reasonable effort. The overly simple or overly complex is boring and ugly, because of the poverty of its perceived patterns. In this paper, I will show that, in consistency with Berlyne's evidence about interestingness and beauty, the scientist does not seek simplicity, per se, but parsimony: that is, pattern in the phenomena.

It is not easy to decide when scientists are interested in parsimony for its own sake (i.e., for its beauty), or when they seek it because the same patterns they find beautiful also are the patterns that describe and explain, and allow them to design and predict. One can speculate that the human emotional response to the beauty of parsimony has evolved through natural selection because it is useful for survival to be able to detect pattern in nature. Certainly the search for pattern in the environment is a very basic and persistent human drive.
Patterns in Data

At the outset, it will be useful to define a few of the terms I have been using, especially simplicity, parsimony and pattern. I am not going to proceed in a highly formal fashion, but will refer to some formalisms, mostly derived from information theory, that have been discussed extensively in the literature. In particular, my discussion is based on the early work of R. J. Solomonoff, 1964 (See also Barron and Cover, 1991). For the rest, I will count on the reader to generalize my definitions of the key concepts to the more complex cases. For a formal account of these matters, generally consistent with what I shall say here, and an excellent review of the literature, see Keuzenkamp and McAleer (1995).

Definitions of simplicity, parsimony and pattern

Simplicity. Consider a data set represented as a one-dimensional ordered string of 0's and 1's. By the complexity of this string, I will mean simply its length, and by its simplicity, the reciprocal of its complexity.¹ Now suppose that there is a formula, also encoded as a string of 0's and 1's (possibly with the addition of a small finite number of special symbols), that can be used to describe with precision the data set. The formula might take the form of a program for generating the successive members of the string. For example, if * is the symbol for indefinite repetition, then the formula (01)* would represent the particular data set that is an alternation of 0's and 1's, and (1)* the data set that is a string of 1's. I can measure the complexity and simplicity of the formula in the same way as

¹On occasion, I will use the term "simplicity" in other ways, but I will always indicate when I am departing from the definition just given.
the complexity and simplicity of the data set — by the formula's length and
the reciprocal of the length, respectively.

Notice that we are concerned with measuring only the complexity of
finite strings, representing observed data sets or formulae we have
generated. There is no occasion to deal with the behavior of the measures
as the sets and formulae grow without limit, although notice from the
examples above that a finite formula can represent a data set of arbitrary
length. The scientist is interested in how well a formula fits a set of data.
Whether the formula will continue to fit as new data are obtained is a
separate question, of interest for theory verification (or falsification), but
irrelevant to discovery. Hence, by restricting our attention to finite sets,
we avoid some of the difficulties that plague most theories of induction.
We will return to this issue from time to time.

**Parsimony and Pattern.** Parsimony is a relation between two
strings: one representing a data set, the other representing a formula for
that set. In general, we will be interested in data sets represented as
sequences of raw observations, before they have been recoded to take
advantage of any redundancy they may possess. The function of formulas
is to exploit such redundancy when it can be discovered.

We could view the formula based on observed pattern as simply a
recoding of the data — a different representation of the same data. We
could then talk of finding the pattern in the data. Alternatively, we can
view the pattern as separate from the data, but denoting it. We will
employ the second viewpoint here, for it is convenient and customary to
distinguish between data and a theory of the data, and viewing the pattern
as simply a recoding of the data confounds the two. Moreover, by keeping pattern and data separate, it is easier to talk about what part of the data are captured by any given pattern, and also to use multiple patterns to capture different components of the data (e.g., to distinguish between the gravitational term and the term for air resistance in a law of falling bodies).

Specifically, *parsimony* is the ratio of the complexity of the data set to the complexity of the formula. To the extent to which a data set can be represented parsimoniously, we say that it is patterned, and we call the representing formula a *pattern*. We will not always insist that the formula represent a data set exactly; in fact, in the real world of science, such representations are almost inevitably approximate (does one need the "almost"?). We can speak of the parsimony of the formula's representation of the data set *up to such and such an approximation*. Except for occasions when we are concerned specifically with goodness of approximation, we will leave implicit the closeness with which the formula describes the data. Whether explicit or implicit, a criterion of goodness of fit is always present in the background of the processes of discovering or testing laws.

Consider two data sets, the second of which includes the first, and is thereby the more complex. Suppose that the same formula describes both sets. Then the parsimony of the relation of the formula with the larger data set is greater than the parsimony of its relation with the smaller set. Similarly, if two formulae describe the same data set, the parsimony of the relation with the simpler formula is greater than the parsimony of the relation with the more complex. As we are most often concerned with comparing the fit of two laws to the same data or the same law to an
expanded data set (new observations), our comparisons will generally only
depend on ordinal properties of the complexity measures, as long as
goodness-of-fit issues do not arise.

Whatever the scientist's motives, an invariable characteristic of laws
is that they simplify the data they describe or explain. We cannot find
laws in data unless there are patterns in the data. If there are patterns,
the data are to that extent redundant and the patterns may be used to
provide a parsimonious description that is more concise than the data
themselves. In this sense, the search for laws is a search for simplicity,
where simplicity here mean terseness of description relative to the original
representation of the data.

The primordial acts of science are to observe phenomena, to seek
patterns (redundancy) in them, and to redescribe them in terms of the
discovered patterns, thereby removing redundancy. The simplicity that is
sought and found beautiful is the simplicity of parsimony, which rests, in
turn, on the exploitation of redundancy. We do not seek the absolutely
simplest law but the law that is simplest in relation to the range of
phenomena it explains, that is most parsimonious.

The Platonist in Search of Pattern

I can recount five anecdotes from my personal experience that
illustrate my Platonic urge to search for pattern. The first three date back
to high school mathematics, where I was delighted to discover from
examples that \((x-y)(x+y) = x^2 - y^2\). I wasn't able, at first, to prove the
relation in general, but I could show empirically that it worked for any
example I tried. It summed up an infinity of facts in a simple formula, which I found beautiful.

A little later in algebra, I found an ugly fact: quadratic equations could have two solutions, one solution or none. I was only reconciled to that apparent lack of pattern when I later learned about imaginary and complex numbers, and that all quadratic equations had exactly two solutions in this extended number system.

Similarly, it was ugly, I thought, that a set of \( n \) linear equations in \( n \) variables, where at least \( k \) of these variables appeared in every subset of \( k \) equations, might or might not have a solution, depending on the ranks of certain determinants. Some years later I was delighted to discover that, if an appropriate measure is defined over the sets of coefficients of the equations, all such equations that do not have a solution lie in a subspace of measure zero (Hall, 1934). (Needless to say, this subspace can be defined by the ranks of the same determinants, but the "rarity" of the exceptions somehow consoled me to their existence.) All of these examples show that discovery of pattern plays a critical role in understanding not only empirical data but also data in which the patterns are created tautologically by the laws of mathematics and logic.

My next two examples relate to empirical matters. At the University of Chicago, while studying econometrics with Professor Henry Schultz, a homework problem required me to fit a function to data where the independent variable was the length of young infants ranging from three to twelve months in age, and the dependent variable was their weight. My Platonist urges led me to fit the function \( y = Ax^3 \), on the hypothesis that
an infant's density and shape would remain approximately constant over this period. Schultz graded the paper B, noting that I should have fitted \( y = Ax^B \), and then tested the significance of the difference, \( B - 3 \). I am still uncertain whether the criticism was justified; after fitting the cubic function I could simply have computed what percentage of the variance in the data was explained by the fitted curve.

On another occasion, Professor Schultz called my attention to A. J. Lotka's wonderful book, *Elements of Physical Biology*, full of fascinating and often mysterious facts. A fact that especially struck my fancy was that if we rank the cities of the United States (or many other countries) by size, we find that a city's population varies (to a very close approximation) with \( 1/R \), where \( R \) is its rank. So the eightieth largest city in the U. S. is almost exactly one tenth as large as the eighth largest city. (In recent decades the rule fits best for metropolitan areas.)

I worked off and on for a number of years before I found an appropriate formula to fit these data (the so-called Yule distribution, which is approximated by the Pareto distribution), and an explanation of why it worked that satisfied me (Simon, 1955). Patterns essentially the same as the rank-size pattern for cities also appear in many other social and biological phenomena: word frequencies in texts, frequency of authorship, sizes of business firms (I'll return to this example later), sizes of biological genera in terms of numbers of species in each, etcetera. It turned out that the (statistical) explanation of the Yule distribution, with a little reinterpretation of the underlying probability mechanisms, also worked for these other phenomena, thus greatly extending the range of application of the pattern without loss of simplicity.
Some Notable Examples of Pattern in Science

A little reading in the history of science assured me that I was not alone in my passion for pattern in data and the parsimony that could be obtained from it. Let me cite a few examples of notable cases where the pattern was found with essentially no help from any pre-existing theory — cases of pure empiricism. One case already mentioned is Kepler's Third Law, which was probably (the evidence is scant) discovered by Kepler in a wholly inductive way, and which, as we will see, has been rediscovered by a computer program using a path that is demonstrably inductive (or abductive, if we prefer Peirce's term for it), receiving no help from theory (Langley, et al., 1987).

A second case is Prout's law (announced about 1815) asserting that all atomic weights are integral multiples of the atomic weight of hydrogen (See Prout's Hypothesis, in Dictionary of the History of Science.) The observed facts were that many atomic weights (for example, all but two of the first twenty), were close to fitting the law although there were egregious exceptions (e.g., the atomic weight of chlorine is 35.5). Chemists took sides on the reality of Prout's pattern and Stoney tested statistically whether the large number of near-integers could have arisen by chance (the odds were very long against it). The issue was settled (in Prout's favor) about a century later with the discovery by Aston of isotopes, and the hypothesis that all atomic nuclei are made up of protons and neutrons with nearly equal masses. The correct law, taking isotopes into account,
was a little less simple than Prout's first approximation, but everyone would now prefer the more exact one.  

A third case is Bode's Law (actually first suggested by Titus in 1766), that the successive planets, Mercury = 0, Venus = 1, . . . , i, . . . are at (approximate) distances \( D_i = 4 + 3 \cdot 2^{(i-1)} \) from the Sun (see Bode's Law in Dictionary of the History of Science). As the "law" is only very approximate, does not hold for the two outer planets, Neptune and Pluto, and has found no clear theoretical explanation, it remains an open question whether there is a law here or a coincidence. A very similar regularity does hold, however, for the satellites of Jupiter, and astrophysicists still seek an explanation, perhaps based on the dynamics of the planetary origins and separation from the Sun.

Notice that in the case of Prout's Law, what tipped the balance to acceptance was a new pattern (involving isotopes) that provided not only a much better fit to the data, but also an explanation — something that is still missing in the case of Bode's Law. The explanation also made more acceptable the remaining deviations, which were very small compared with those that had been removed, hence might be expected to require a different explanation (a different pattern or mechanism) to remove them. Today, we attribute them to the nuclear packing fraction.

A fourth case, the last I shall mention, is Balmer's law (1885) for the wave lengths of successive lines in the hydrogen spectrum: \( \frac{i^2}{(i^2-4)}, i = 3, \)

---

2 Even the revised law is not exact, for it does not take care of the packing fractions, whose explanation depends on the mass-energy equivalence introduced by special relativity. Hence physicists today might prefer for some purposes a second approximation by a still more complex law.

January 6, 1997 - 11 - Science Seeks Parsimony
4, ... ... (soon generalized to $i^2/(i^2-j^2)$) (Banet, 1966). It fits the data with extraordinary precision, was discovered by a non-physicist by pure induction — perhaps "numerology" would better describe the process — and had no explanation until 1913 when Bohr provided one with his new quantum model of the hydrogen atom. Here the fit was so good and the parsimony so great that the law obtained immediate acceptance in 1885 even without an explanation.

So finding pattern is the name of the game — or at least of one very important game in science. I mention again that we are speaking both of exact patterns discovered inductively in mathematical "data" (my high school examples) and approximate patterns in real data. In a few cases, e.g., Balmer's Law, the pattern may fit the data extremely well; often, the fit is quite approximate. Later, I will return to the question of approximation, and how we should judge whether an approximation is satisfactory or unsatisfactory.

The Uses of Parsimony

We see that parsimony is at the root of what we mean by a scientific law. With simplicity, it also plays several roles in the discovery of laws and their verification. Parsimony is not only the end product of scientific activity, but it also helps to guide the discovery and verification processes. Let us see how this comes about.
Discovery versus Verification of Laws

Philosophers of science commonly distinguish between the discovery of laws and their verification (Reichenbach, 1938). Early writers, including Sir Francis Bacon, in *The New Organon*, John Stuart Mill, in *A System of Logic*, and William Whewell, in *The Philosophy of Inductive Sciences* paid considerable attention to law discovery, and laid down normative principles to guide efficient experimentation. Consider, for example, John Mill's Canon of Difference:

If an instance in which a phenomenon occurs and one in which it does not differ in only one other circumstance, it is the cause or the effect, or an indispensable part of the cause, of the phenomenon.

If we observe a phenomenon to occur from time to time but not always, we can seek to determine its causes and effects by searching for the circumstances that distinguish its appearance from its nonappearance. If we find such a circumstance, we can never be sure that it is unique, but that need not discourage us, for we are not looking for an infallible method of scientific discovery — something that no one has promised us. Discovery is a process of heuristic search whose success cannot be guaranteed.

Because the search for laws is a heuristic process carrying no guarantees of success or validity, many philosophers of science in this century, perhaps the most notable being Karl Popper\(^3\), denied the

---

\(^3\)He calls on the authority of, among others, Albert Einstein, who is quoted (*The Logic of Scientific Discovery*, p. 32) as saying: that "there is no logical path leading to
possibility of formulating normative rules of discovery. Only during the past 35 years, beginning with Norwood Hanson's *Patterns of Discovery*, has this view been seriously challenged. However, with the construction in recent years of a number of computer programs that incorporate successful mechanisms of discovery, we possess a rather conclusive constructive demonstration that normative (heuristic) theories of discovery are possible (Langley, et al., 1987).

We will consider separately the roles of simplicity and parsimony in discovering (generating) theories and their roles in verifying or falsifying (testing) them. Then we will say something also about simplicity in the application of theory and simplicity in experimentation. We will see that simplicity is not always used in the sense we defined earlier, but has somewhat different meanings and different roles in each of these activities.

**Parsimony and Simplicity in Law Discovery**

We may take as common, though not universal, the case of some observed phenomena in need of description and explanation: that is, which call for a law that fits them — if possible, a law representing a mechanism that explains them. There may or may not already exist some theory that could play the role of this law. As a first sub-case, let us assume that no such theory exists. We are then confronted with the task of data-driven law discovery.
**Data-Driven Discovery.** The BACON program (Langley, et al., 1987) has shown itself to be capable of discovering laws by fitting functions to data. At the most general level, BACON consists of a hypothesis generator and a test. The hypothesis generator produces functions, the test fits them to the data and decides that the fit is or is not satisfactory. (The programmer sets the standard of approximation that must be met.) Now the easiest way in which to build such a hypothesis generator is to begin with a small set of simple functions, and generate successively more complex ones from combinations of these primitives. The next function to be generated is selected on the basis of information about the fit or misfit of the functions already generated.

As the functions are not simple binary strings, hence have no obvious unique order, our earlier definition of simplicity does not apply without extension. One meaning of simplicity in this case is just the order of generation. We call a function simple if it is generated early in the sequence. This order may be different for different sets of data for, as remarked above, observed relations with the data of the functions already generated affect what function will be generated next. As a function is complex (in an informal sense) to the extent that it is formed as some kind of combination of previously generated functions, the new definition of simplicity is not wholly unrelated to the basic one for binary strings.

Consider again the specific example of Kepler's Third Law. The given data (we can use Kepler's actual data) are the distances from the Sun ($x$) and periods of revolution ($y$) of the then-known planets. BACON observes that as $x$ grows larger $y$ also grows larger, and forms the hypothesis that $y/x = k$, a constant. When this function fails to fit the data,
BACON observes that as $x$ grows larger $y/x$ also grows larger and hypothesizes that $x/(y/x) = x^2/y = k$, a constant. The fit is now closer, but the test fails again.\footnote{As in all such procedures in science, the closeness of fit that is regarded as acceptable is an arbitrary parameter. Accepting or rejecting Prout's Law (or any other) is a matter of taste -- or should we call it free will?} Persevering, BACON next observes that as $y/x$ grows larger $x^2/y$ grows larger, and hypothesizes that $(y/x)/(x^2/y) = (y^2/x^3) = k$, a constant. In this case, the test succeeds, yielding Kepler's Third Law: a planet's period of revolution varies as the $3/2$ power of its distance.

Now it seems quite natural to build the BACON system from a small set of primitives that themselves are simple in another sense: simple in having only a few parameters. Just as Peano's Axioms generate all the integers from zero with the help of the single operation of succession, so BACON generates a large set of functions from the linear function with the help of the two operations of multiplication and division, applied in response to cues provided by the data. By moving from the simple to the complex (in this new sense) we increase the likelihood that we will find first the function that provides greatest parsimony.

It is of some interest with respect to the trade-off between simplicity and parsimony that Kepler, ten years before he found the $3/2$-power law, had published the hypothesis that the planets' period varied as the square of the distance. Only some years later did he become dissatisfied with the fit of hypothesis to data and search for a law that gave better agreement. It happens that BACON's generator also discovers the square law before the $3/2$-power law. Whether it accepts the former, or rejects it and goes on to the latter, depends on the goodness-of-fit criterion it applies (determined by the programmer). BACON, like Kepler and Prout (and

\begin{footnotesize}
\begin{itemize}
\item[4] As in all such procedures in science, the closeness of fit that is regarded as acceptable is an arbitrary parameter. Accepting or rejecting Prout's Law (or any other) is a matter of taste -- or should we call it free will?
\end{itemize}
\end{footnotesize}
everyone else), needs a separate parameter (a "propensity for simplicity") to determine what degree of approximation is acceptable.

For a wider range of applications in science, we might want to supply BACON with a few more primitives (e.g., the log function, the exponential, the sine function), but it is remarkable that with the linear function as its sole primitive, it discovers not only Kepler's Third Law, but also Joseph Black's law of the equilibrium temperatures of mixtures of liquids, Ohm's law of current and resistance, Snell's law of refraction, the law of conservation of momentum and a host of others. What is more remarkable, it achieves each of these successes after only a small number of trials. What it definitely doesn't do is to "try out all possible functions." Instead, its search is highly selective and guided by cues.

In the course of discovering laws that fit data, BACON also discovers a number of theoretical concepts, including inertial mass, specific heat, voltage, molecular weight, atomic weight (distinguishing the latter two) and others. The motivation for these discoveries of theoretical concepts is that, by introducing them, laws previously found can be simplified, hence can explain the same data more parsimoniously (Simon, 1985).

For example, BACON finds that when bodies A and B are accelerated by a stretched spring connecting them, the ratio of their accelerations is always the same constant $k_{AB}$. Bodies A and C provide another constant, $k_{AC}$, and bodies B and C a third, $k_{BC}$. BACON, finding a new pattern, that $k_{AB}$ times $k_{BC}$ equals $k_{AC}$, postulates a constant associated with each body, $m_A$, $m_B$, and $m_C$, respectively, and hypothesizes that the ratio of the accelerations of I and J, $k_{IJ} = m_j/m_i$, for all I and J. If we have N objects,
we can now state the law relating the accelerations of all pairs of them in terms of N constants, the $m_i$, instead of $N^2$ constants, the $k_{ij}$. The m's are, of course, familiar to us as inertial masses. By introducing theoretical terms we have gained a new source of parsimony.

From this brief (and somewhat oversimplified) description of BACON, we see at least three roles for simplicity in law discovery or generation. First, as hypotheses must be tried in some order, the easiest way to build a hypothesis-generator of some generality is to start with a simple primitive function, or a small set of simple primitive functions, and generate successively more complex functions by the application of simple combinatorial operations (in our example, multiplication and division).

The second role of simplicity is to make as small as possible the number of functions that have to be generated before one is found that fits the data. In discovery systems like BACON, the order of function generation is not fixed, but responds to the form of the misfit in unsuccessful applications. This feedback serves as a powerful heuristic that often permits satisfactory functions to be found with very little search.

The third role of simplicity is to make more parsimonious the laws that are found, by introducing new theoretical terms (new properties that are not directly observable but are inferred from observables). Thus, important theoretical terms in physics (e.g., mass, specific heat, voltage and the like) can be discovered in response to the goal of simplifying the forms of laws (Langley et al., 1987).
In all of these roles, simplicity goes hand in hand with parsimony. That is to say, we seek to explain the data with simple functions, containing few constant parameters, before we test complex functions. Similarly, in introducing theoretical terms, we greatly reduce the number of independent parameters whose values must be fixed to explain the data. This is the aspect of parsimony, that Popper, with his emphasis on verification, defended. In fact, he argued (p. 140) that "[t]he epistemological questions which arise in connection with the concept of simplicity can all be answered if we equate this concept with degree of falsifiability [his italics]." Popper did not, however, see that the concern for parsimony can lead to normative rules for discovery systems: that such systems should be designed, as far as possible, to generate simple rules before generating complex ones.

The progression in hypothesis generation from simple to complex is not peculiar to the BACON program but is seen in most of the other extant examples of law-finding programs. For example, Valdes-Perez's MECHEM (1992) accepts data on the substances that are input to a chemical reaction and some of the substances that are produced by it, and then attempts to find a sequence of reaction steps that will account quantitatively for these inputs and outputs.

MECHEM does this by first trying all possible one-step reactions (1) with no unobserved intermediate products, (2) that are consistent with applicable constraints derived from chemical theory. Then, if it fails to match the data, it gradually increases the number of steps and the number of postulated unobserved reactants until a reaction chain is found that fits the data. In this respect MECHEM is like theorem-proving programs that
first search for one-step proofs, then for two-step proofs, building each new structure on those previously generated. The architecture that proceeds from simplicity to complexity in the generation of hypotheses discovers the maximum redundancy in the data and is easily implemented.

**Theory-Guided Discovery.** The examples discussed thus far, except MECHEM, deal with discovery when no theory, or essentially no theory, exists to guide the search for pattern. This is, in fact, a common case in the history of science, but certainly not universal. Often existing theory, while not adequate to account of the phenomena of interest, can provide powerful help in the search. Consider the question of finding a law that gives the equilibrium temperature of the mixture of a liquid contained in two vessels, each with its own initial mass and temperature (Black's Law). If it is hypothesized that the law must be symmetrical with respect to the two liquids (a hypothesis that itself contributes to simplicity of the law) then, after the form of the function for the first liquid is determined, the same form can be postulated for the second. Likewise, if it is hypothesized that the mass and the heat (the product of temperature by mass) will be conserved, this additional constraint will further limit the set of functions to be generated.

In fact, when BACON is given these three constraints, of symmetry and mass and heat conservation, it finds the law without any superfluous search whatsoever. In this case, in fact, it deduces the law from previously accepted theory, just as Newton deduced Kepler's Third Law from his own laws of mechanics. Later on, we will have occasion to look at other cases of law discovery where the relation between prior theories and new patterns is more complex than in these two examples.
Parsimony and Simplicity in Verification

I have already mentioned the important role of simplicity in enhancing parsimony, thereby making a theory easier to falsify. Of course we know today, as a result of the work of Lakatos and others, that the matter is not quite as simple as Popper made it. What we usually falsify by our observations or experiments is not a single law but the consistency with the evidence of a whole complex of laws, together with assumed initial and boundary conditions. We can usually save any particular part of this complex by sacrificing some other part. I will not go into the whole Lakatosian story of progressive and degenerating research programs, but simply, for the present, ignore this added complication and speak of falsification in a simple Popperian sense.

Direct Comparison of Theory with Data. Each degree of freedom — each parameter in our theory that must be estimated from the set of data that is being used to test the theory — both adds to the complexity of the theory and weakens our ability to falsify it if it is wrong. Consider Ohm's Law, I = kV/R, where I is the current, V the voltage, R the resistance, and k a constant. In Ohm's original experiment, he could measure the resistance (the length of a wire) and the current (the angle of the needle on an ammeter), but not the voltage. Thus, he "used up" one set of observations to estimate the product, kV, and this single trial could not falsify the hypothesis. Only after estimating kV could he then test his theory by modifying R (changing the length of the resistance wire) and inserting in the equations the values observed in a second experimental trial (with kV held constant by using the same battery).
In Ohm's experiment, the problem of fitting parameters is very simple, whereas the usual situation in econometrics (or in most observational, as distinguished from experimental, sciences) is far more difficult, not in principle but quantitatively, because we often want to consider the simultaneous interaction of many variables. Suppose we wish to fit $N$ data points in $V$ variables by a system of simultaneous linear equations. As we must estimate $V^2 + V$ parameters from the data, we will always obtain a perfect fit unless $N > V^2 + V$, hence only in the latter case can we test the hypothesis that the data fit the model.

If we wish to experiment with functional forms that have more parameters, the situation gets progressively worse. If we have a sufficient number of data points so that the model can be falsified, we can then introduce an error term into each equation, apply the method of least squares and use the percent of variance in the data that is explained to determine whether we wish to accept the equations as describing the pattern of the phenomena. The underlying rationale is essentially the same as BACON's. As in all statistical testing, the criterion of acceptance is arbitrary.

Identification of Structural Relations. It would take me too far afield in this essay to discuss the whole concept of identification and the reasons why, both in order to understand mechanisms and in order to apply theories to practical affairs, sciences always strongly prefer identified structural equations to equations that are incompletely identified (Koopmans, 1949). The relation of identifiability to parsimony may be briefly characterized as follows. Suppose we have a theory that postulates $k$ mechanisms governing the behavior of $k$ variables. We now
come to a situation where we have reason to believe that \( k-I \) of the mechanisms remain effective, but the remaining one has been altered. For reasons of parsimony, we would prefer to carry over the unaltered parts of the theory, making modifications only in the equation corresponding to the mechanism that has been changed. If we have information about the nature of the change, we may even be able to use this information to infer the modified mechanism. We do not have to recreate the theory, whenever it has suffered a structural change, from whole cloth.

Structural equation systems are parsimonious descriptions of reality, factoring complex systems into simpler subsystems. The approximate effect of successive factorization, when it is possible, is to reduce the complexity of a system from a linear to a logarithmic function of the number of its elements.

To the extent that we can match the components of the equation systems to corresponding components of the systems they describe, we can build theories of complex systems in combinatorial fashion from the theories of their components. The existing body of theory in molecular biology is a striking example of the payoff that can be obtained from this divide-and-conquer strategy. Factorability is an powerful source of simplicity, and parsimony. (See the discussion of near-decomposability in the last pages of this paper.)

**Matching Inferences from Theory to Data.** We turn now to another verification issue. In the situations examined so far, we have been concerned with matching laws directly to data. However, in many situations, we must draw out the implications of theories before we can
test them against available data. For example, Einstein had to reason from his theory of time and synchronization of clocks to the Lorentz transformations and from these to the invariance of the Maxwell's equations under Lorentz transformations in order to test special relativity (assuming that Maxwell's equations had already been found to agree with the data on electromagnetism). To take a simpler example, Newton had to deduce Kepler's laws from the laws of motion and the inverse square law of gravitational attraction in order to show that the latter laws were consistent with the planetary motions.

**Simplicity in Computation.** In these inference processes, another kind of simplicity may be important: simplicity that facilitates computation. Drawing out the implications of a theory requires deduction or simulation or some other form of computation, which may be simple or complex. If a set of differential equations can be solved in closed form, the paths that the equations define can be compared directly with data to test their fit. In the literature of theoretical mechanics, we find many papers that solve a system of simultaneous equations for the stresses in a board with, say, an infinite slit or other unrealistic boundary conditions. In these cases, the boundary conditions are not selected because there are many real boards in the world with infinite slits, but because, if these boundary conditions, rather than more realistic ones, are used, the equations can be solved in closed form.

Even in our present world of supercomputers, it is still a trivial matter to find problems of deep scientific interest where heroic simplification is required to make feasible the computations needed to understand the behavior of the system of interest — either by solving in
closed form or by simulation. The recent focus of attention on chaotic systems has greatly increased our sensitivity to computational feasibility. The computer has reduced but not eliminated the need for simplification and approximation as practical necessities for testing and applying theory.

The field of computer science has been much occupied with questions of computational complexity, the obverse of computational simplicity. But in the literature of the field, "complexity" usually means something quite different from my meaning of it in the present context. Largely for reasons of mathematical attainability, and at the expense of relevance, theorems of computational complexity have mainly addressed worst-case behavior of computational algorithms as the size of the data set grows larger. In the limit, they have even focused on computability in the sense of Gödel, and Turing and the halting problem. I must confess that these concerns produce in me a great feeling of ennui.

When I speak here of computational complexity (or simplicity), I refer to the amount of computation we have to do to solve a problem of some kind in the average case, and what percentage of the problems of that kind we are likely, on the basis of previous experience, to solve with an acceptable amount of computation ("acceptable" being a flexible word like "significant"). Just as in the long run we are all dead, so, as problems

---

I have been startled, in the past several years, to find that the phrase "bounded rationality," for which I can claim some authorship, has now been borrowed by rational expectationists (e.g., Sargent) and game theorists (e.g., Aumann) to refer to worst-case complexity theory and the limits on computation imposed by Gödel and Turing. Bless me, the actual bounds on the powers of computation that we mortals possess are so severe that we almost never have occasion, in our scientific practice, to regret that we are only Turing machines and cursed with Gödel incompleteness! Our difficulties with computation start at much lower levels, not even including the worst case. But then, I belong to a generation that remembers how matrices were inverted with hand-operated desk calculators.
grow in size without limit, we don't solve them. Scientists do not ask for guarantees (and certainly not guarantees in the form of theorems) that they will solve any given problem, or even that the problem has a solution (a parsimonious formula for the data). They follow various procedures in search of patterns (like the BACON procedures described earlier) and sometimes they are successful — often enough to maintain their addiction to science. (Experimental psychology has shown that random partial reinforcement is more conducive to addiction than consistent reinforcement.)

Finally, the advantages of simplicity (in my sense of the term) for computation must be balanced against the loss of fit of theory to data through approximation. I will take up this issue at some length after considering other implications of simplicity for the application of theories.

**Simplicity in the Application of Science**

From our discussion of computational simplicity, we can see that simplicity plays a role in applying accepted theories that is similar to the role it plays in testing new theories. Parsimony enhances our ability to make strong predictions about the behavior of the systems of interest, whether they be systems we are designing or natural systems, by decreasing the number of degrees of freedom we lose through parameter estimation. At the same time, computational simplicity increases our ability to deduce or to simulate system behavior under the anticipated conditions of application.

Notice that parsimony and computational simplicity are not always compatible. When we limit ourselves to functional forms that facilitate
computation, we may have to introduce additional parameters in order to obtain sufficiently close approximations to the real phenomena. For example, Fourier series and Taylor's series provide powerful formalisms for approximating wide classes of functions, but at the expense of introducing a potentially infinite number of parameters. These formalisms afford computational simplicity but not necessarily parsimony. The problem thereby created is serious if our goal is to test a theory but not if we are simply applying an accepted theory whose validity we are not questioning. Even in the latter case, lack of parsimony makes it impossible for us to make strong predictions from the theory when we have to estimate most or all of the parameters from the data.

Simplicity in Experimentation

I will distinguish two classes of experiments (or observational procedures), one of which may be called well-structured, the other exploratory. They call for quite different heuristics in their design.

Well-Structured Experiments. The design of experiments, as usually described, provides an obvious application of yet another kind of simplicity. According to the standard doctrine, we vary one variable at a time, controlling other variables that are known to be or suspected of being influential, and we randomize where we cannot control. We seek situations where the variables we are measuring will sound loud and clear against the background of unwanted noise. Strong structure and controls of this kind are of special value when we already have a specific hypothesis, originating from a theory or elsewhere, that we want to test.
Then a good experimental design clears away irrelevancies and focuses attention on the variables of interest.

**Exploratory Procedures.** Well-structured experiments fit well the Popperian picture of science in which the main concern is the verification or falsification of hypotheses that already exist. The generation of hypotheses is a different matter. Let us distinguish two cases. In one case, we already have a general theory. We deduce some new consequences from the theory and then design an experiment to test whether these consequences are supported empirically. This is the pattern of what Kuhn called "normal science." In the other case, there are some phenomena of interest — bacteria, say — but at the moment we have no hypothesis about them that we wish to test. Then we can design experiments, or merely opportunities for observation, that may reveal new and unexpected patterns to us. Having found such patterns, we can search for explanations of them.

Most of my earlier examples of law discovery fit this latter description of scientific activity. In my algebra classes I noticed certain patterns in equations: when the sum and difference of two quantities were multiplied, the product always had the same form: the difference of their squares. A quadratic function usually had two solutions (but there were exceptions), and n linear equations in n variables generally had a unique solution (but again there were exceptions). Prout found lots of elements whose atomic weights were integral multiples of the atomic weight of hydrogen, an improbable result if only chance were at work. Balmer found a regularity in the spectral lines of hydrogen. Kepler found a relation between the distances and periods of revolution of the planets.
Let us consider two examples of a somewhat more complex sort. Fleming observed, in a Petri dish that he had left unwashed in his laboratory when he had gone on vacation, some bacteria being lysed, and near them, a growth of fungus that he recognized as belonging to the genus *Penicillium*. You and I (unless you are a bacteriologist) would have noticed nothing, for we would see nothing unusual in bacteria cells that were dissolving in the proximity of a fungus (or even *that* it was a fungus). As Pasteur observed, accidents happen to the prepared mind.

Fleming's observation led him to a very vague hypothesis: that there might be a cause and effect relation between the presence of the fungus and the fate of the bacteria. He (like many other scientists) had a standard way (a heuristic) for approaching unexpected patterns when only vague hypotheses (or none) were available: (1) to design experiments to test the scope of the phenomena, and (2) to look for a mechanism capable of producing them. Fleming's (incomplete) follow-up pursuing this heuristic, and its later completion by Florey and Chain, brought the discovery of antibiotics.

The second example concerns the young Hans Krebs, who took as his first major independent research project designing a sequence of experiments to elucidate the chemical reaction path that produces urea in living mammals (Holmes, 1991). He did not have a strong hypothesis but a question (and some research tools he thought might help find the answer). He did know (a weak hypothesis) that the likely sources of the nitrogen in urea were ammonia and the amino acids. However, the task he set himself was not to test that hypothesis but to accept it and design experiments that might disclose the mechanism (the reaction path).
Consequently, he added various mixtures of ammonia and particular amino acids to slices of living liver tissue and measured the products. In many such experiments, he observed only a minimal production of urea, but on an occasion when he tested ammonia together with the amino acid ornithine, he obtained a large yield. For his prepared mind, and on the basis of his recent experiments, this was a contrasting pattern and a surprising one. He responded by doing exactly what Fleming did: (1) he tested the scope of the phenomenon ((a) did it require ornithine, or would similar substances do the trick?; (b) what happened to the product when the quantities of ammonia and ornithine were varied?), (2) he searched for a plausible chemical reaction path. Within a few months he had found the path for urea synthesis.

The cases of Fleming and Krebs are not unusual in the history of science. One could easily produce another dozen that led to Nobel Prizes. A person who is expert in some domain exposes himself or herself to phenomena, watches for the unexpected (a pattern), then designs new observations or experiments to account for it. One attractive feature of this procedure is that it is tolerant of complexity in the phenomena being observed. All that is required is that these phenomena have a variety of properties (a pattern) which someone skilled in the art will come to know. Then a "surprising" pattern is just one that is different from the familiar pattern.

A simpler example of pattern finding helps to show what is going on here. Standard intelligence tests often make use of the Thurstone Letter Series Completion task (Simon and Kotovsky (1963). A series of numbers or letters is presented, say, ABMCDMEF..., and the test-taker is asked to
supply the continuation (in this case, MGHM would be the "right" answer).

A strong case can be made for this task as a good measure of aptitude for high level cognitive skills. Studies of how answers are found reveal a very simple picture. Subjects notice two kinds of relations, repetitions of the same symbol and relations of next between pairs of symbols, where "next" refers to succession on some familiar "alphabet." For example, speakers of English in our culture would find relations of next between X and Y, Tuesday and Wednesday, October and November, the integers 16 and 17, and so on.

Having noticed such relations in the test series, the test-taker finds a simple pattern, usually involving periodic repetition and hierarchies of the relations, that accounts for the presented sequence and allows it to be extrapolated. A very small number of primitive relations, those we usually associate with the concept of symmetry, account for the major patterns that have been found in sciences ranging from molecular genetics to particle physics. The pattern that is discovered is formed combinatorially from the small set of primitive relations.

Which Form of Experiment? In observing the natural world we do not have these luxuries of varying things one at a time and removing noise; and therein lies the superior power of laboratory experiments for disclosing the underlying structural relations in which we are interested, once we know what variables we should segregate and examine. Therein also lies the weakness of formal experimentation when it is designed tightly to test particular hypotheses: it does not usually provide us with the surprises that often lead to our most exciting and important scientific discoveries. For these — for the initial hypotheses that guide our
experiments — we must depend largely on observation of complex phenomena against the background of a large body of stored expert knowledge, with the aim of encountering surprises.

Exploration, as distinct from systematic experimentation, is not regarded very favorably in typical normative theories of experimental design. "Random" search in complex situations is thought to be a highly inefficient procedure. On the contrary, for a scientist who has built up a wide body of knowledge about some domain, any situation that conflicts with that knowledge, hence violates expectations, is a starting point for a program of experimentation, using the "surprise" heuristic described earlier in the cases of Fleming and Krebs as a basis for guiding search.

Rapid scanning of the environment until an unexpected pattern, a surprise, occurs has proved historically to have been a highly productive scientific strategy (not to the exclusion, of course, of other strategies). A plausible account in terms of the surprise mechanism can be given for many instances of the "aha" phenomena that are so greatly treasured by historians of science, and thought so mysterious by the layperson.

**Simplicity versus Closeness of Fit**

In building theories, there is almost always a trade-off between simplicity and close approximation to the facts. My anecdote about the "theory" relating lengths and weights of infants, \( W = aL^3 \), illustrates the point well. By assuming that density and shape do not change with growth, we save one degree of freedom in our equation. Should we make this assumption or should we weaken the theory by substituting a parameter, \( b \), for the number 3? Or should we further decompose \( b \) into a
product of two parameters, one for change in density, the other for change in shape? The obvious answer is that "it depends." Depends on what? It depends on how close our assumptions approximate the reality and on how accurate a theory we need or want.

The Principle of "Unrealism"

In economics, Milton Friedman's advocacy of the "principle of unrealism" (Friedman, 1953) has provided a frequent battleground for the contention between simplicity and closeness of approximation. One of his examples, Galileo's law of uniform acceleration, celebrates Galileo's genius in sticking with simplicity (parsimony in relation to the data). For understanding the movement of bodies in space where air resistance is negligible, or even close to Earth at low velocities, extracting the relation between a constant force and a velocity that grows steadily with time \( v = gt \) was a stellar scientific achievement. It also provides an excellent illustration of the experimental strategy of varying one variable at a time in order to determine the underlying mechanisms that are present in, and interact in, more complex situations. In fact, its value for law discovery and perhaps for verification is far more obvious than its value for application of the law in either design or prediction.

For example, Galileo's discovery fails to advance the science far enough to support the design and manufacture of parachutes. The maker of parachutes might prefer, to \( v = gt \) or its equivalent \( \frac{dv}{dt} = g \), a formula that looked more like \( \frac{dv}{dt} = (g - fv) \), (leading to the even more complex \( v = \frac{g}{f}(1-e^{-ft}) \)) where \( f \) is a parameter representing the air resistance on the body. True, there is another parameter here to be estimated, and it
will take on different values for bodies of different sizes and shapes, and air of different densities. But the new formula has the virtue that, these estimates having been made, it might predict the behavior of parachutes (e.g., their terminal velocity, \( g/f \)), which the simpler formula can not. Hence, as the two formulae have different ranges of application, it is meaningless to ask which is the more parsimonious. The first thing we must ask of a formula, before considering its parsimony, is whether it describes the data of interest to an acceptable level of approximation.

**Simplicity and Complexity in Modern Biology**

Before turning to examples drawn from economic theory, let me discuss one more example from the natural sciences: simplicity and approximation in genetics (Watson, 1976). We start with Mendel, who in his experiments with sweet peas, arrives at the important concept of the gene and a simple probabilistic law of inheritance for combinations of genes that (by good luck in his choice of experimental material) are inherited independently and exhibit a full dominant/recessive contrast. Mendel was able to explain his data by means of an extremely simple theory (which is still retained as a special case in the modern genetic model) as a result of the lucky fact that the particular situations he investigated had these particular special properties.

Before the end of the 19th Century, chromosomes were observed under the microscope and were identified as the agents of inheritance because of their behavior (splitting) during meiosis. These observations produced new data that, interpreted within the existing model, produced a great gain in parsimony.
The next major steps occurred at the turn of the 20th Century when, along with the rediscovery of Mendel's work and its association with chromosomes, crossovers of chromosomes were observed (accounting for statistical dependence between genes). The observation of chromosomes had strengthened the theory (by providing a hypothetic mechanism for the observed phenomena), but certainly complicated it. The discovery of crossover complicated it further by introducing the parameters that described the dependence between pairs of genes, but also provided a potential explanation in terms of actual location of the genes on the same or different chromosomes. Here we have a classic case of the tradeoffs among simplicity of hypotheses, range of phenomena explained and falsifiability.

The idea that genes could be assigned particular locations on particular chromosomes suggested new experimental manipulations — for example, increasing mutation frequency by irradiation — and led to curiosity about the chemical structure of genes and chromosomes. Nearly a half century passed before answers began to be found for these questions: first, identifying DNA rather than proteins as the basis for differentiation among chromosomes (which contained both); and second, identifying protein synthesis as the key process governed by genes, through the mapping of nucleotide triads upon amino acids. The introduction of these mechanisms produced a far more complex theory (which became still more complex as myriads of details were discovered), but a theory that could begin to explain the entire process in terms of accepted principles of physical chemistry, hence was actually enormously parsimonious.
During the entire century to which I am referring, there appeared to be no reluctance whatsoever to moving from simpler theories to more complex ones that gave far more complete explanations (and predictions) of the phenomena. Parsimony was valued highly, as was a detailed understanding of complex phenomena, but bare simplicity of theory, which could only be obtained by sacrificing explanatory power, was valued hardly at all.

Laws of Qualitative Structure

Along with the formal genetic models that have been developed, with their exquisitely detailed picture of the chemical processes, there have also emerged some broad qualitative generalizations of great importance — what Allen Newell and I, in another context (Newell and Simon, 1976), called "Laws of Qualitative Structure." In genetics, an example of such a law is: "A protein is a sequence of amino acids that is (often) synthesized in vivo on a template consisting of a linear sequence of nucleotides (RNA), the latter mapped 1-1 from a corresponding sequence (DNA) in the chromosome." This is the meaning of the slogan: "One gene, one protein."

Notice the "usually," and "often" in the statement of the law, as well as its general vagueness. Many of the most important laws of biology (and of physics, for that matter) have this qualitative and slogan-like character: for examples, the germ theory of disease, and the cell theory. The germ theory of disease simply amounts to the advice that: "If you encounter a disease, look for a micro-organism as cause; there may often be one. Of course there are many diseases that don't involve micro-organisms." The
cell theory suggests that: "Organisms are divided into cells each of which has many organelles, including (at least in eukaryotes) a nucleus which contains (most of) the genetic material, and various organelles capable of carrying out metabolic processes."

While these laws of qualitative structure are no substitute for detailed and relatively rigorous theories, they play an important role in organizing the thinking of experts in scientific domains about the phenomena with which they deal, and in guiding their search for solutions of the problems they attack. For example, the germ theory of disease suggests the heuristic of applying to an ailing organism the known methods for detecting and identifying microorganisms in tissue. One might say that they provide simple, but very rough, statements of very complex but powerful theories. Scientists carry both the "slogans" and the detailed theories around, and alternate between using the one or the other as circumstances dictate. I will have more to say later about what these circumstances might be.

**Parsimony and Simplicity in Economics**

The same principles of theory building and choices between simplicity and complexity that I have been illustrating with examples drawn from physics and biology apply, with little need for modification, to economics. Again, concrete examples (which I have drawn from my own work) will help us see why (Simon, 1997).
Utility Maximization v. Bounded Rationality

The maximization of expected utility, the central axiom of neoclassical economics, is usually regarded as an exceedingly simple theory, because all that it postulates is a consistency of choice behavior over the whole set of choices that an actor makes. Given the subject's utility function, income and probability distributions over future events, and the full set of market prices, his or her demands for all goods are determined. Income and the market prices can be observed, hence create no problem; but as the utility function is completely undetermined apart from the requirement of consistency, virtually any pattern of demands is consistent with the theory; and in spite of "revealed preference," there is generally no practical way to describe empirically the content of the utility function, hence the actual pattern of demand, at any given moment of time.

Matters are a little better if we compare two situations that are close in time, the second being identical to the first except that the price of one commodity has increased. Now, with a few additional "harmless" assumptions about the utility function (essentially that it is unchanged and that there are no inferior commodities) we can predict that less of the commodity whose price has risen will be demanded than before. However, without yet another assumption — or information — (about elasticities), we cannot predict whether the total amount spent on that commodity will increase or decrease, much less what the magnitude of the change will be. Thus, each bit of additional prediction seems to call for empirical evidence
about one or more additional parameters. The theory may be thought simple, but it is not especially parsimonious.

**Measuring Demand Elasticities.** Henry Schultz, in his monumental *The Theory and Measurement of Demand*, did what scientists in other fields usually do in such circumstances: he attempted to determine the parameters, in particular the elasticities of demand, empirically. In the process he revealed the difficult issues of identification of structural equations that face such an undertaking. What was further discouraging was that the parameters, if estimated, could hardly be expected to retain their values over any great length of time. Measuring demand elasticities was more like measuring the magnetic field of the Earth, or even the changing patterns of barometric pressures, than it was like measuring the velocity of light, the gravitational constant, or Planck's constant.

Moreover, it was not clear what should be taken as the independent variables. The theory of choice under uncertainty was in a primitive state at that time, and although Schultz was well aware that he had to make assumptions about the predictive process, the assumptions he chose looked much more like the cobweb than like rational expectations. (There was, of course, no empirical evidence available in the 1930's — and there is precious little today — that would have led very readily to the choice of either of these models or some other.)

**Marginal Utilities and Budgets.** Let me turn to my own first research encounter with marginal utilities. In 1935, I studied at first hand the administration of public recreation programs in my native city of
Among other things, I was interested in the budget process that required choices between planting more trees on playgrounds, say, and hiring additional recreation leaders. Obviously (or so my economics courses had taught me) the money should be put where the marginal return was greatest. There were only a few difficulties in applying this principle: (1) you had to be able to estimate the results produced by the alternative expenditures, and (2) you had to be able to determine the increments in utility produced by these results.

As a practical matter, my observations and interviews showed that none of the participants in the process were thinking of the matter in this way, as they had not the slightest idea of how to carry out the requisite measurements. The observations showed further that, in the two departments involved in operating the program, administrators affiliated with the city's public works department almost always preferred the first alternative (trees), whereas administrators affiliated with the school board almost always preferred the second (recreation leaders).

Nevertheless, budget decisions were made. In the process, the principal references to data were references to last year's expenditures for these same purposes. If it was possible to spend more in the coming year, the increment was usually allocated more or less proportionately to previous spending levels. Of course the process was more complex than this (at least more wordy), but these remarks convey its flavor. Out of my observations (and the accompanying surprise at the irrelevance of what I had learned in Henry Simons' excellent course in price theory), emerged some laws of qualitative structure dealing with budget processes and human decision-making processes in general.
Bounded Rationality and Organizational Loyalties. The first of these new laws of qualitative structure was the principle of bounded rationality: Human beings have reasons for what they do, but they seldom maximize utility. They do not maximize because, given the complexities and uncertainties of the choice situations they face, they have neither the knowledge (theories and facts) nor the computational abilities to carry out the necessary inference processes. The principle of bounded rationality is, like the germ theory of disease (and like utility theory, for that matter), a scientist's hunting license rather than a definite answer to the question of what behavior will be observed. It's simplicity is deceptive, for applying it requires one to ascertain a myriad of facts.

Fortunately (and in contrast to utility theory), forty years of research in cognitive psychology within the information processing framework has disclosed many of these facts, especially the mechanisms that are used in decision making when maximization of SEU is impossible. The resulting theory is not simple (see, for example, Chapter 14 of Newell and Simon, 1972), but it is parsimonious, allowing prediction of behavior in some detail and with some accuracy in a large number of situations that have been studied, including some within the domain of economics. It looks much more like theory in molecular biology than like Newton's Laws of Motion.

The second law of qualitative structure that came out of my recreation study was that members of organizations identify, cognitively and motivationally, with the goals of the organizations or organization units to which they belong. This principle also has central importance to economics: specifically (1) to the question of the nature of altruism and the
role that altruism plays in human choice, and (2) to the motivations that underlie behavior in organizations and the consequences of these motivations for the respective roles of organizations and markets in an economy. (Simon 1990, 1991).

Both laws of qualitative structure that emerged from my first adventure in research had some of the same seductive simplicity that is possessed by SEU maximization. As in the latter case, the attempt to apply them revealed what a vast amount of empirical information must be provided before broad principles of this kind (or even more rigorous ones that are expressible in equations) can make predictions about specific situations.

Explanation of Supply and Demand Changes. Before leaving this issue, let us look at one more example, which I think not atypical, of an application of SEU maximization theory: Gary Becker's "explanation" of the movement of women after WWII into the labor market (Becker, 1981). Clearly, the SEU theory requires that something change either in women's utility functions (and hence in the supply function) or in the demand for their labor. Becker opts for the latter without providing any empirical evidence for his choice or even discussing the possibility of changes in the utility function.

Surely it is not implausible to suppose that, as a result of the gradual change of women's role in society during the first half of this century, smaller families and the experience of women in the workplace during World War II, there have been large continuing changes in their preferences for outside employment versus managing a household. Clearly
the question of whether the increase in employment resulted from a shift in demand or supply (hence utility) is an empirical one and the theory, without such evidence, does not support one explanation over the other.

The "work" of arriving at Becker's explanation is not done by the theory but by unsupported empirical assumptions which could only be validated by data — data that were not provided. Even more damaging to Becker's claims, the same phenomenon of increasing employment of women could have resulted from mechanisms consistent with bounded rationality, without appeal to the SEU maximization hypothesis. Just to suggest two from among numerous possibilities, the wartime experience may have called women's attention to the possibilities of outside employment, of which they had been only dimly aware, or may have redefined what constituted a satisfactory pattern of life activities.

Are we better or worse off with respect to simplicity or parsimony if we substitute bounded rationality for utility maximization to explain this and other concrete phenomena? We have seen that the principle of bounded rationality is clearly a law of qualitative structure, subject to many kinds of imprecision. It advises us that, in order to predict behavior, we must know something not only about the preferences of the actor and the environment in which he or she is acting (which utility theory also requires) but also about the actor's knowledge (of alternatives and of consequences) and ability to compute and draw inferences.

What does bounded rationality propose as a replacement for utility maximization? It says that we must examine human behavior empirically to find out what procedures are actually used to make decisions. These
procedures may vary considerably, depending on what knowledge and information is available and upon the complexity of the situation. They may change in the course of history with the acquisition of new knowledge and computational techniques, and with shifts in public attention. The procedures that are used in decision are not constructed arbitrarily and anew, however, for each decision situation, but themselves derive from more general principles.

The psychological research I mentioned earlier has brought together and tested some much more specific models of the decision process, which incorporate such mechanisms as heuristic search (including search to enlarge the commodity space), satisficing and aspiration levels (finding alternatives that meet "satisfactory" levels, where what is satisfactory is constantly modified by experience), means-ends analysis, focus of attention, recognition, the content and organization of expert memories, and so on.

We might say that this body of theory resembles, qualitatively if not quantitatively, theory in molecular biology. It postulates a large number of specific mechanisms, and a great deal of empirical knowledge is required before it can be applied in any particular circumstances. Neither SEU maximization nor bounded rationality are exceptionally parsimonious theories. Nor are they simple, except when they are posed in the forms of Laws of Qualitative Structure — rough qualitative guides to the phenomena. How are we to choose between them?

I would propose that we should choose in much the same way that we choose in other sciences. First, we should seek empirical evidence of
whether the mechanisms advanced by each theory do in fact operate as postulated. To the extent that they do, the theory at least has explanatory value, if not predictive value. Of course, we might conclude that there is no way of finding out about these mechanisms. To that extent, we would have to regard the theory as vacuous: "metaphysical" or "non-operational" in positivist language.

If our interests are more applied than basic, we might choose between theories, as Friedman proposes, on the basis of predictive power rather than explanatory value. But if we do that, we must include not just the bare bones of the theory but the parameters that have to be estimated in order to make the predictions. Here again, a theory is essentially helpless unless there are practical means for estimating these parameters, and the success of the prediction is apt to rest at least as heavily on the parameter measurements as on the exact form of the axioms.

**Firm Size Distributions.** I turn to another example from my own economic research to illustrate the points I have just made. In the chapters on the theory of the firm in neoclassical economics textbooks, some attention is usually given to the determinants of firm size. The standard theory transmitted by all the textbooks with which I am familiar stems from Jacob Viner and his celebrated debate with Wang, his draftsman. Each firm, according to Viner as amended by Wang, has a U-shaped short-term cost curve, determined by the level of operation for which it was designed, and a U-shaped long-term cost curve which is the envelope of the short-run curves for all possible levels of operation. In the long run, firms will operate on a scale that corresponds to the point of minimum cost on the long-term cost curve.
What does this theory (deduced from profit maximization) allow one to say about the distribution of firms by size in an industry or in an entire economy? Essentially nothing unless all the firms in the industry face the same long-term cost curve. In that case, all firms will be predicted to be about the same size, a prediction that is in egregious conflict with all of the known facts. If, on the other hand, each firm has an idiosyncratic cost curve, then each will have its size determined by that curve. In this case, nothing at all can be said about the size distribution of firms in the industry without previous knowledge of the optimum for each firm — not a very parsimonious prediction.

As I mentioned earlier in this paper, I came to the firm size question from the opposite direction: from empirical data of actual size distributions, which almost always fit quite well the Yule (or Pareto) distribution, a distribution related to, but even more asymmetrical than, the lognormal distribution (Ijiri and Simon, 1977). For the phenomena of firm sizes, this distribution can be derived from assumptions that make very much weaker demands on human calculation than does SEU maximization, and that are compatible with human bounded rationality. The basic assumption (usually called the Gibrat assumption) is that the average rates of growth of firms of all sizes are independent of present size. If, on average, small firms grow by five per cent per year, than medium-sized firms also grow five per cent on average, and large firms, five per cent.

This Gibrat assumption has several things to commend it. First it is simple — much simpler than SEU maximization, hence far more parsimonious. Second, it is easy to find data to test it. Third, and perhaps
most important, it fits the facts very well in almost all cases where it has been tested. Fourth, the Gibrat assumption itself follows from some rather plausible assumptions about the underlying mechanisms: roughly, that the amount of money a firm can borrow at reasonable cost for expansion, the speed with which it can expand its production and marketing facilities and markets at reasonable cost, and all of the similar magnitudes that govern overall expansion will be roughly proportionate to the present size of the firm. (If we want to put the matter in terms of U-shaped cost curves, we should begin with curves for the cost of changing the rate of production, rather than curves for the cost of production at different output levels.) I don't know that these hypotheses about ease of expansion have ever been tested systematically, but such casual data as I am familiar with fits them, and they "resonate" when proposed to people experienced in business.

In the case of business firm sizes, it seems clear that the principles usually applied in science to choose among theories — whether simplicity, parsimony, or predictive power and accuracy — will pick the explanation of firm size distributions based on the Gibrat assumption over the nearly vacuous explanation based on profit maximization. That the Viner-Wang theory persists in the textbooks, which seldom mention the Gibrat theory, can only be explained by the mystique of SEU that continues to dominate the profession of economics. It is as if physicists continued to believe that space was occupied by an ether for forty years after the results of the Michelson-Morley experiments had been published.
Parsimony and Simplicity in the World

Up to this point, we have been concerned with the meaning of simplicity and parsimony in scientific theories. In this final section, still focusing on my own research, I would like to propose some generalizations about simplicity in the phenomena that we study. I will examine two examples — one relating to the human sciences, the other more abstract and general. The lessons I will draw are, first, that simplicity (as distinguished from parsimony) is attainable in science to the extent that the phenomena we are studying are simple; and second, that, fortunately for us, we find in many domains a particular kind of simplicity that enables us to pursue a divide-and-conquer strategy in our research.

Is Human Behavior Simple?

In the discussion thus far of simplicity and parsimony in science, I have made no distinction between the human sciences and the so-called "natural" sciences. Some special considerations arise when we are building theories about the behavior of adaptive organisms, among which human beings are of particular interest to us. Our investigations will lead to some generalizations about the relation between the complexity of the environments in which adaptive systems live and the simplicity or complexity of the systems themselves (and hence of theories about them).

In another context (Simon, 1996) I claim that, if we set aside for a moment the contents of human memory, then "[a] man, viewed as a behaving system, is quite simple. The apparent complexity of his behavior
over time is largely a reflection of the complexity of the environment in which he finds himself." The reason for exempting memory from this characterization is that the contents of memory largely reflect the complexity (or simplicity) of the environment. They are not part of the built-in human architecture, but are acquired and shaped through experience of the world (including, of course, social experience).

Here we have another law of qualitative structure. What does it assert about the simplicity or complexity of theory? It asserts that if we divide our theory of human behavior into two parts — (1) architecture, and (2) learned knowledge and skill — the former will be relatively simple (at least at the level of detail of information processes, although not necessarily at the neuronal level) in comparison with the latter. The learning processes themselves, being part of the architecture, will be simple, but the body of knowledge and skill acquired through them will not be. Contemporary theories of cognition, including theories of problem solving and of perception, learning and memory, partition matters in this way, acquiring thereby enormous parsimony. Let me comment on one such theory which has been employed to explain, among other things, the bases for specialized human expertise.

EPAM is a model of human perception, learning, memory (short-term and long-term) and recognition. The theory was first proposed by Feigenbaum (1961) to account for the phenomena of human rote verbal learning, and in the succeeding third of a century has been gradually extended to explain a wide range of phenomena that have been studied by experimental psychologists. In recent years, EPAM has been of particular value for explaining the vast differences between the performances of
novices and experts in any particular domain (Richman, et al., 1996). To accomplish this, it interacts with another theory, the General Problem Solver (GPS), whose history dates back to 1958, and which provides an explanation of human problem solving processes (Newell and Simon, 1972). EPAM and GPS, taken together, may be regarded as specifications at the symbolic level of the main mechanisms that implement bounded rationality in human beings.

I will not attempt to describe either EPAM or GPS, but will simply show how they provide a simple theory of expertise — one that is simple at least at the level of qualitative structure, and not too complex even when we look at the nitty-gritty details of implementation. However we assess its simplicity, the theory is surely parsimonious — probably the more important assessment — for EPAM and GPS use only a few basic mechanisms to account for exceedingly complex behavior (e.g., playing grandmaster chess, making medical diagnoses, discovering scientific laws, steering an automobile, writing music, making drawings).

The central mechanisms of GPS are heuristic search and, in particular, means-ends analysis, which I have already discussed. The central mechanisms of EPAM are a learning process and a recognition process. Through the learning process EPAM acquires the ability to distinguish among stimuli that are presented to it. It does this by using very simple processes to build a discrimination net of unlimited size (nets of more than 70,000 terminal nodes have been grown) that performs successive tests on stimulus features and sorts them accordingly. The discrimination process enables EPAM to recognize stimuli already
familiarized through learning, and thereby to access information about them that has been stored previously.

In brief, the theory of expertise asserts that experts are able to perform as they do because (1) by virtue of GPS-like structures and processes, they have powerful heuristics that makes their problem solving in the domain of expertise orders of magnitude more powerful and efficient than the problem solving of novices; (2) by virtue of EPAM-like structures and processes, they recognize a great many situations from the cues they perceive in them, and thereby gain access to knowledge and skills relevant to these situations that enable them to solve many problems by immediate recognition and that greatly enhance their search efficiency in solving the others. One might say that the expert operates like a well-indexed encyclopedia, EPAM serving as index, with capabilities (GPS) for selective search.

The evidence supporting this account of expertise has been gathered in many domains and is by now overwhelming. One can say with assurance, for example, that a great deal of medical diagnosis is achieved by recognition (of symptoms), and when this does not suffice, the rest of the job is done by heuristic search that guides the collection of additional evidence and its interpretation. The theory has not lacked application to economic phenomena, elucidating, among others, the processes used by a financial analyst to determine a firm's problem areas, the processes used to select an investment portfolio, systems for scheduling complex phenomena when optimization is not computationally feasible, job choice processes, the processes used to analyse a business case of the kind commonly used in business policy courses, and the investment decision
processes of business firms (see references to this research in Simon, 1997). Recently, it has scored some major successes in showing how bounded rationality (it might in this case be called minimal rationality) can bring about relatively efficient equilibration of markets in the absence of mechanisms for utility maximization.

By constructing a domain-independent theory of heuristic search (GPS) and recognition (EPAM) mechanisms, and by supplying domain knowledge that has been determined independently to be available to experts in specific domains, theories in the form of running computer models have been able to account for expert performance in these domains. Because the computer models can themselves perform professional tasks in the domain, they make direct predictions of expert behavior that can be compared with observations of human experts in the laboratory or in real-life settings. As the examples of the previous paragraph show, these methods can be used to study and explain, in a way that can be empirically verified or refuted, the decision-making of economic actors.

Are Large Systems Simple?

The final topic that I shall discuss had its beginnings in economics, but has important implications for such distant domains as evolutionary biology and computer science, and indeed, for the organization of the sciences in general. It concerns the relation between particular kinds of pattern in the phenomena under study, on the one hand, and the parsimony of the theories that describe the phenomena, on the other. It
goes under the labels of "near decomposability" and "nearly complete decomposability" of dynamic systems, and the core of it is found in the Simon-Ando theorem, first published in *Econometrica* in 1961.

**Block Triangular and Diagonal Dynamic Systems.** Shortly after its publication, my attention was attracted to a paper by Richard M. Goodwin (1947). Goodwin was concerned with the justification of partial equilibrium analysis, especially the legitimacy of treating sets of variables as unconnected if their connection is very weak or if they are coupled unilaterally (systems with unilateral coupling of subsets of variables are called block triangular).

Exploring this idea further, I was led to a formal characterization of causal ordering and its relation to identifiability (Simon, 1953). In 1957, Albert Ando and I, returning again to Goodwin's ideas, saw how to simplify dramatically, without serious loss of information, the description of systems that were nearly completely decomposable, and in this way to make their descriptions much more parsimonious. Nearly completely decomposable systems are systems made up of blocks of variables, where each block is only weakly connected with the others (Simon and Ando, 1961). For example, in the input-output matrices of economic systems, it is typical for variables within particular clusters of industries to be tightly bound with one another, but only weakly bound with the variables of other clusters. The variables and equations of nearly completely decomposable systems can be arranged so that their matrices are nearly block diagonal (Table 1).
It occurred to us that this hierarchical structure of relations among variables and sets of variables was not peculiar to economic systems or human organizations; it was also commonly observed in both inorganic nature (layers from quarks through atoms to molecules) and organic nature (layers from organic molecules through complex reactions to cells, organs and organisms).

![Diagram](image)

Table 1
A Nearly Completely Decomposable Matrix

X represents a possibly strong link;
\( \varepsilon \), a weak (epsilon) link

This led us to the hypothesis (a law of qualitative structure) that systems in nature that appear to be complex will usually be found to be patterned in hierarchies; and that they will be describable simply and parsimoniously if advantage is taken of this special and powerful property.
We were able to prove (the Simon-Ando Theorem) that in a stable dynamic system whose matrix of coefficients is nearly block diagonal (Table 1), each of the diagonal blocks rapidly approaches its steady state almost independently of the others. Over a longer period of time, the entire matrix gradually approaches a long-term steady state; and during this second period the diagonal blocks move monolithically, nearly preserving their individual steady states. Thus, to describe the entire system approximately, it is necessary only to solve the equations for each of the diagonal blocks, replace each block by a single element that is a weighted average of the elements of its principal eigenvector, and then solve this aggregated system (with one degree of freedom for each block) for the long-term behavior of the system.

**Multi-Level Systems.** This possibility of simplifying nearly completely decomposable systems through aggregation is not limited to systems of two levels, like the one depicted in Figure 1. Each of the diagonal blocks in that figure could itself represent a nearly completely decomposable system, and so on, to any number of levels. In terms of time scales, events at the lowest level of the hierarchy would occur very rapidly, then on slower and slower time scales as we ascend toward higher levels. Aggregation amounts to selecting out the processes that operate on each particular time scale and studying their behavior independently of the faster and slower processes. For the time scale selected, dynamic phenomena are observed at a particular level of aggregation; variables at the higher levels of aggregation remain approximately constant; and

---

*Physicists, struggling with certain problems in quantum electrodynamics appear to have rediscovered a closely similar idea, which they employ under the label of "renormalization."*
subsystems at lower levels are seen only as aggregates in steady-state motion.

**Near-decomposability and the Specialization of Science.** In a world built on this plan, as large aspects of our world are, phenomena on different time scales can be studied nearly independently of each other; and a good deal of the specialization among the sciences is based precisely on this possibility for decomposition and aggregation. Each of the sciences selects phenomena on a particular time scale, treats variables on slower time scales as constant parameters, and treats subsystems operating on faster time scales as aggregates whose internal details are irrelevant at the scale of special interest. This partitioning is perhaps seen in its sharpest form in physics and chemistry, in the progression from quarks, through the so-called elementary particles, atomic components, to small molecules, macro-molecules, etcetera.

Near-decomposability is of great importance for the social sciences also. For example, the cognitive theories that I have described in this paper postulate processes that have durations ranging from tens of milliseconds to a few seconds. (It may take about ten milliseconds to test a single feature in a stimulus; about 500 milliseconds to recognize a stimulus; about 8 seconds to store a simple new stimulus in long-term memory.) At the next level below are neural processes that are one or more orders of magnitude faster. (It takes about one millisecond for a signal to cross the synapse between two neurons.) It is this difference between the speeds of processes at the two levels that permits us to build a theory of cognition at the symbol-processing level which is reducible, in principle, to a neuro-psychological theory, but which can be developed
independently of the details of the latter theory. To understand problem solving, we don't need to know much detail about the operation of neurons.

A second way of representing the near-decomposability of complex systems is by means of computer programs that employ closed subroutines. In such programs, each routine is executed by executing in sequence its component routines, until we come down to the level of primitive instructions, which are executed directly (translated directly into machine language). A routine will execute correctly as long as it provides the correct inputs to its subroutines, and the subroutines, in return, produce correct outputs. Beyond this requirement, the operation of a routine is wholly independent of the structure and operation of its subroutines (and of their sub-subroutines, down to the level of the primitives). Just as markets reduce the information that economic actors must have about the other parts of the economic system, so the architecture of closed subroutines, and the closely related architecture of nearly complete decomposability, reduce the dependence of each component on the detailed structure and operation of the others.

"Natural" Simplicity in Theories. The suspension of theories from hierarchical sky-hooks is characteristic of all of science. One might describe it as Nature's concession to the bounded rationality of scientists: phenomena do not have to be understood all at once, but can be divided into components that can be studied separately. But the subdividing process, if it is to be successful, must respect the structure of the natural systems. The hierarchical structure of many complex systems, and the divide-and-conquer strategy it enables, is a major source of simplicity in
scientific theories just because it is a major source of pattern. The independent treatment of system components does not represent simplicity for simplicity's sake, but the exploitation of the parsimony that is present in nature. In particular, near-decomposability gains this parsimony with little loss in accuracy of approximation. The simplicity that is achieved is a by-product of the way things are, not a celebration of a "principle of unrealism."

**Evolutionary Basis of Near-Decomposability.** It is natural to ask how we are so fortunate as to find nature structured in this way. A plausible answer can be given in terms of natural selection. As I have given that answer at considerable length in *The Sciences of the Artificial* (1996), I will only outline it briefly here. Although it has not been shown conclusively, there are reasons to believe that the rate at which complex systems are likely to evolve is very much accelerated if they are composed of stable subsystems, each of which can be altered without major alterations to the others. At any level of complexity (measured, say by system size) hierarchically ordered systems will evolve more rapidly than systems that are not so ordered. As a result, most of the very complex systems that have evolved, and that we are able to observe in the world, have hierarchical structure.

In inorganic nature, and organic nature up to the level of single-celled organisms, the advantage of near-decomposition is gained whenever simpler stable systems combine into more complex stable systems (e.g., atoms into molecules). At the level of multi-celled organisms, similar advantages are obtained because the interrelations of the stable chemical processes of specialized tissues and organs resemble the interrelations of
closed subroutines: each depends only on the outputs of the others and
provides inputs to the others. The success of each is nearly indifferent to
the details of how its inputs are produced by the providers.

The liver converts ammonia previously derived from amino acids
into urea, which it delivers to the kidneys for excretion. Provided that the
urea is synthesized and delivered, the operation of the kidneys is
unaffected by the exact sequence of urea-synthesizing reactions in the
liver. Increased efficiency of the liver will not affect the kidneys in the
short run; over a longer run the ratio of sizes of the two organs may
change if one gains efficiency more rapidly than the other. The important
fact is that the rate of evolution is accelerated by the mutual independence
of subsystems from sensitivity to each other's details of structure and
process.

The implications of near decomposability for efficiency and evolution
of complex systems is still only imperfectly understood. We can expect
that progress in molecular developmental biology will soon throw a great
deal of light on these matters.
Conclusion

What have we learned from this exploration of simplicity and parsimony in science? In particular, what are the principal laws of qualitative structure that we can carry away from our investigation? I will not attempt a detailed summary of the many issues that I have discussed, but simply point to some of the most important of the generalizations that emerged from the discussion.

The basic desideratum in science, whether pursued for understanding, for its practical products or for its beauty, is parsimony, which may be measured by the ratio of the number of symbols required to describe the data individually to the number required to describe the patterns they form. We aim to discover pattern in observed facts that can be used to describe and explain these facts parsimoniously.

Parsimony brings simplicity in its wake; but simplicity in theory without parsimony in the relation between the theory and data is bought only at the price of weakening the goodness of approximation of our descriptions, narrowing the range of phenomena over which they extend, and impoverishing our understanding of the phenomena.

Law discovery requires us to synthesize hypotheses, and it is usually advantageous to synthesize simple hypotheses before complex ones, both (1) because it is easiest to generate hypotheses by combinatoric operations on a few primitive hypotheses, and (2) because we are seeking the simplest hypothesis that will account for the phenomena (that is, the most parsimonious).
Parsimony measured by the ratio of number of data points to number of parameters that need to be estimated from the data, enhances the falsifiability of theories.

Simplicity, when it can be attained without unacceptable sacrifice of descriptive, explanatory and predictive power, can facilitate computation with theories and reasoning about them. Computational simplicity, simplicity of statement of theories, and simplicity of the mechanisms that implement theories are not always synonymous. Computational simplicity is a principal grounds for choosing between alternative representations of the same theory.

We must distinguish between the simplicity of a theory and the simplicity of the mechanisms at the next level down, implied by the theory. However one regards the simplicity of SEU theory, its implementation implies unattainably complex computational mechanisms for the theory's human actors. On the other hand, bounded rationality can be implemented by computational mechanisms that are within human capabilities.

Describing data in terms of theories that are built on structural relations provides means for predicting the behavior of systems under partial change in structure. The validity of the structural relations in a theory can only be tested if the theory is overidentified.

Important features of theories are often encapsulated in laws of qualitative structure, which are heuristic and not rigorous, and which
provide high-level generalizations, and representations useful in organizing problem-solving search.

The source of simplicity in a theory is the existence of discoverable pattern in the phenomena the theory describes and explains. The hierarchical structure of many of the phenomena of our world provides a basis for constructing independent, and relatively simple, theories for different problem domains, each dealing with system behavior at a different level of the hierarchy. The levels are distinguishable in terms of the speeds and durations of their processes, and nearly-independent theories can be provided for the phenomena at each hierarchical level. Among large systems, evolution will select favorably for those that are hierarchical.

References


