# 6William C. Wimsatt

# Robustness, Reliability, and Overdetermination (1981)

(from M. Brewer and B. Collins, eds., (1981); *Scientific Inquiry in the Social Science*s (a festschrift for Donald T. Campbell), San Francisco: Jossey-Bass, pp. 123-162.)

> Philosophy ought to imitate the successful sciences in its methods, so far as to proceed only from tangible premises which can be subjected to careful scrutiny, and to trust rather to the multitude and variety of its arguments than to the conclusiveness of any one. Its reasoning should not form a chain which is no stronger than its weakest link, but a cable whose fibers may be so slender, provided they are sufficiently numerous and intimately connected [Peirce, [1868] 1936, p. 141] .

> Our truth is the intersection of independent lies [Levins, 1966, p. 423].

The use of multiple means of determination to "triangulate" on the existence and character of a common phenomenon, object, or result has had a long tradition in science but has seldom been a matter of primary focus. As with many traditions, it is traceable to Aristotle, who valued having multiple explanations of a phenomenon, and it may also be involved in his distinction between special objects of sense and common sensibles. It is implicit though not emphasized in the distinction between primary and secondary qualities from Galileo onward. It is arguably one of several conceptions involved in Whewell's method of the "consilience of inductions" (Laudan, 1971) and is to be found in several places in Peirce.

Indeed, it is to be found widely among the writings of various scientists and philosophers but, remarkably, seems almost invariably to be relegated to footnotes, parenthetical remarks, or suggestive paragraphs that appear without warning and vanish without further issue. While I will point to a number of different applications of multiple determination which have surfaced in the literature, Donald Campbell has done far more than anyone else to make multiple determination a central focus of his work and to draw a variety of methodological, ontological, and epistemological conclusions from its use (see Campbell, 1958, 1966, 1969a, 1977; Campbell and Fiske, 1959; Cook and Campbell, 1979). This theme is as important a contribution as his work on evolutionary epistemology; indeed, it must be a major unappreciated component of the latter:

multiple determination, because of its implications for increasing reliability, is a fundamental and universal feature of sophisticated organic design and functional organization and can be expected wherever selection processes are to be found.

Multiple determination—or *robustness,* as I will call it—is not limited in its relevance to evolutionary contexts, however. Because of its multiplicity of uses, it is implicit in a variety of criteria, problem-solving procedures, and cognitive heuristics which have been widely used by scientists in different fields, and is rich in still insufficiently studied methodological and philosophical implications. Some of these I will discuss, some I will only mention, but each contains fruitful directions for future research.

## Common Features of Concepts of Robustness

The family of criteria and procedures which I seek to describe in their various uses might be called *robustness analysis.* They all involve the following procedures:

1. To analyze a *variety* of *independent* derivation, identification, or measurement processes.
2. To look for and analyze things which are *invariant* over or *identical* in the conclusions or results of these processes.
3. To determine the *scope* of the processes across which they are invariant and the *conditions* on which their invariance depends.
4. To analyze and explain any relevant *failures of invariance.*

I will call things which are invariant under this analysis "robust," extending the usage of Levins (1966, p. 423), who first introduced me to the term and idea and who, after Campbell, has probably contributed most to its analysis (see Levins,1966,1968).

These features are expressed in very general terms, as they must be to cover the wide variety of different practices and procedures to which they apply. Thus, the different processes in clause 1 and the invariances in clause 2 may refer in different cases to any of the following:

a. Using different sensory modalities to detect the same property or entity (in the latter case by the detection of spatiotemporal boundaries which are relatively invariant across different sensory modalities) (Campbell, 1958, 1966).
b. Using different experimental procedures to verify the same empirical relationships or generate the same phenomenon (Campbell and Fiske, 1959).
c. Using different assumptions, models, or axiomatizations to derive the same result or theorem (Feynman, 1965; Levins, 1966; Glymour, 1980).
d. Using the agreement of different tests, scales, or indices for different traits, as measured by different methods, in ordering a set of entities as a criterion for the "validity" (or reality) of the constructed property (or "construct") in terms of which the orderings of entities agree (Cronbach and Meehl, 1955; Campbell and Fiske, 1959).

e. Discovering invariance of a macrostate description, variable, law, or regularity over different sets of microstate conditions, and also determining the microstate conditions under which these invariances may fail to hold (Levins, 1966, 1968; Wimsatt, 1976a, 1976b, 1980b).

f. Using matches and mismatches between theoretical descriptions of the same phenomenon or system at different levels of organization, together with Leibniz's law (basically that if two things are identical, no mismatches are allowed), to generate new hypotheses and to modify and refine the theories at one or more of the levels (Wimsatt, 1976a, 1976b, 1979).

g. Using failures of invariance or matching in a through f above to calibrate or recalibrate our measuring apparatus (for a, b. or f) or tests (for d), or to establish conditions (and limitations on them) under which the invariance holds or may be expected to fail, and (for all of the above) to use this information to guide the search for explanations as to why the invariances should hold or fail (Campbell, 1966, 1969a; Wimsatt, 1976a, 1976b).

h. Using matches or mismatches in different determinations of the value of theoretical parameters to test and confirm or infirm component hypotheses of a complex theory (Glymour, 1980) and, in a formally analogous manner, to test and localize faults in integrated circuits.

One may ask whether any set of such diverse activities as would fit all these items and as exemplified in the expanded discussion below are usefully combined under the umbrella term *robustness analysis*. I believe that the answer must be yes, for two reasons. First, all the variants and uses of robustness have a common theme in the distinguishing of the real from the illusory; the reliable from the unreliable; the objective from the subjective; the object of focus from artifacts of perspective; and, in general, that which is regarded as ontologically and epistemologically trustworthy and valuable from that which is unreliable, ungeneralizable, worthless, and fleeting. The variations of use of these procedures in different applications introduce different variant tools or consequences which issue from this core theme and are explicable in terms of it. Second, all these procedures require at least partial *independence* of the various processes across which invariance is shown. And each of them is subject to a kind of systematic error leading to a kind of *illusory robustness* when we are led, on less than definitive evidence, to presume independence and our presumption turns out to be incorrect. Thus, a broad class of fallacious inferences in science can be understood and analyzed as a kind of failure of robustness.

Nonetheless, the richness and variety of these procedures require that we go beyond this general categorization to understand robustness. To understand fully the variety of its applications and its central importance to scientific methodology, detailed case studies of robustness analysis are required in each of the areas of science and philosophy where it is used.

## Robustness and the Structure of Theories

In the second of his popular lectures on the character of physical law, Feynman (1965) distinguishes two approaches to the structure of physical theory: the Greek and the Babylonian approaches. The Greek (or Euclidean) approach is the familiar axiomatic one in which the fundamental principles of a science are taken as axioms, from which the rest are derived as theorems. There is an established order of importance, of ontological or epistemological priority, from the axioms out to the farthest theorems. The "Greek" theorist achieves postulational

economy or simplicity by making only a small number of assumptions and deriving the rest — often reducing the assumptions, in the name of simplicity or elegance, to the minimal set necessary to derive the given theorems. The "Babylonian," in contrast, works with an approach that is much less well ordered and sees a theoretical structure that is much more richly connected:

> So the first thing we have to accept is that even in mathematics you can start in different places. If all these various theorems are interconnected by reasoning there is no real way to say "These are the most fundamental axioms," because if you were told something different instead you could also run the reasoning the other way. It is like a bridge with lots of members, and it is overconnected; if pieces have dropped out you can reconnect it another way. The mathematical tradition of today is to start with some particular ideas which are chosen by some kind of convention to be axioms, and then to build up the structure from there. What I have called the Babylonian idea is to say, "I happen to know this, and I happen to know that, and maybe I know that; and I work everything out from there. Tomorrow I may forget that this is true, but remember that something else is true, so I can reconstruct it all again. I am never quite sure of where I am supposed to begin or where I am supposed to end. I just remember enough all the time so that as the memory fades and some of the pieces fall out I can put the thing back together again every day" [Feynman, 1965, pp. 4647].

This rich connectivity has several consequences for the theoretical structure and its components. First, as Feynman (1965, pp. 5455) observes, most of the fundamental laws turn out to be characterizable and derivable in a variety of different ways from a variety of different assumptions: "One of the amazing characteristics of nature is the variety of interpretational schemes which is possible. It turns out that it is only possible because the laws are just so, special and delicate.... If you modify the laws much you find that you can only write them in fewer ways. I always find that mysterious, and I do not understand the reason why it is that the correct laws of physics seem to be expressible in such a tremendous variety of ways. They seem to be able to get through several wickets at the same time." Although Feynman nowhere explicitly says so, his own choice of examples and other considerations that will emerge later suggest another ordering principle for fundamentality among laws of nature: *The more fundamental laws will be those that are independently derivable in a larger number of ways*. I will return to this suggestion later.

Second, Feynman also observes that this multiple derivability of physical laws has its advantages, for it makes the overall structure much less prone to collapse:

> At present we believe that the laws of physics have to have the local character and also the minimum principle, but we do not really know. If you have a structure that is only partly accurate, and something is going to fail, then if you write it with just the right axioms maybe only one axiom fails and the rest remain, you need only change one little thing. But if you write it with another set of axioms they may all collapse, because they all lean on that one thing that fails. We cannot tell ahead of time, without some intuition, which is the best way to write it so that we can find out the new situation. We must always keep all the alternative ways of looking at a thing in our heads; so physicists do

> Babylonian mathematics, and pay but little attention to the precise reasoning from fixed axioms [Feynman, 1965, p. 54].

This multiple derivability not only makes the overall structure more reliable but also has an effect on its individual components. Those components of the structure which are most insulated from change (and thus the most probable foci for continuity through scientific revolutions) are those laws which are most robust and, on the above criterion, most fundamental. This criterion of fundamentality would thus make it natural (though by no means inevitable) that the most fundamental laws would be the least likely to change. *Given that different degrees of robustness ought to confer different degrees of stability, robustness ought to be a promising tool for analyzing scientific change.* Alternatively, the analysis of different degrees of change in different parts of a scientific theory may afford a way of detecting or measuring robustness.

I wish to elaborate and illustrate the force of Feynman's remarks arguing for the Babylonian rather than the Greek or Euclidean approach by some simple considerations suggested by the statistical theory of reliability. (For an excellent review of work in reliability theory, see Barlow and Proschan, 1975, though no one has, to my knowledge, applied it in this context.)

A major rationale for the traditional axiomatic view of science is to see it as an attempt to make the structure of scientific theory as reliable as possible by starting with, as axioms, the minimal number of assumptions which are as certain as possible and operating on them with rules which are as certain as possible (deductive rules which are truth preserving). In the attempt to secure high reliability, the focus is on total elimination of error, not on recognizing that it will occur and on controlling its effects: it is a structure in which, if no errors are introduced in the assumptions and if no errors are made in choosing or in applying the rules, no errors will occur. No effort is spared in the attempt to prevent these kinds of errors from occurring. But it does not follow that this is the best structure for dealing with errors (for example, by minimizing their effects or making them easier to find) if they do occur. In fact, it is not. To see how well it handles errors that do occur, let us try to model the effect of applying the Greek or Euclidian strategy to a real (errorprone) theory constructed and manipulated by real (fallible) operators.

For simplicity, assume that any operation, be it choosing an assumption or applying a rule, has a small but finite probability of error, $p_0$. (In this discussion, I will assume that the probability of error is constant across different components and operations. Qualitatively similar results obtain when it is not.) Consider now the deductive derivation of a theorem requiring $m$ operations. If the probabilities of failure in these operations are independent, then the probability of a successful derivation is just the product of the probabilities of success, $(1 - p_0)$, at each operation. Where $p_s$ stands for the probability of failing at this complex task ($p_s$ because this is a serial task), then we have for the probability of success, $(1 - p_s)$:

$$(1 - p_s) = (1 - p_0)^m$$

No matter how small $p_0$ is, as long as it is finite, longer serial deductions (with larger values of $m$) have monotonically decreasing probabilities of successful completion, approaching zero in the limit. *Fallible thinkers should avoid long serial chains of reasoning.* Indeed, we see here that the common metaphor for deductive reasoning as a chain is a poor one for evaluating probability of failure in reasoning. Chains always fail at their weakest links, chains of reasoning only most probably so.

When a chain fails, the release in tension protects other parts of the chain. As a result, failures in such a chain are not independent, since the occurrence of one failure prevents other failures. In this model, however, we are assuming that failures are independent of each other, and we are talking about probability of failure rather than actual failure. These differences result in a serious disanalogy with the metaphor of the argument as a chain. A chain is only as strong as the weakest link, but it is that strong; and one often hears this metaphor as a rule given for evaluating the reliability of arguments (see, for example, the quote from C. S. Peirce that begins this chapter). But a chain in which failure could occur at any point is always weaker than (in that it has a higher probability of failure than) its weakest link, except if the probability of failure everywhere else goes to zero. This happens when the weakest link in a chain breaks, but not when one link in an argument fails.

Is there any corrective medicine for this cumulative effect on the probability of error, in which small probabilities of error in even very reliable components cumulatively add up to almost inevitable failure? Happily there is. *With independent alternative ways of deriving a result, the result is always surer than its weakest derivation.* (Indeed, it is always surer than its *strongest* derivation.) This mode of organization— with independent alternative modes of operation and success if any one works—is parallel organization, with its probability of failure, $p_p$. Since failure can occur if and only if each of the $m$ independent alternatives fails (assume, again, with identical probabilities $p_0$ ):

$$p_p = p_0^m$$

But $p_0$ is presumably always less than 1; thus, for $m > 1$, $p_p$ is always less than $p_0$. Adding alternatives (or redundancy, as it is often called) always increases reliability, as von Neumann (1956) argued in his classic paper on building reliable automata with unreliable components. Increasing reliability through parallel organization is a fundamental principle of organic design and reliability engineering generally. It works for theories as well as it does for polyploidy, primary metabolism, predator avoidance, microprocessor architecture, Apollo moon shots, test construction, and the structure of juries.

Suppose we start, then, with a Babylonian (or Byzantine?) structure—a multiply connected, poorly ordered scientific theory having no principles singled out as axioms, containing many different ways of getting to a given conclusion and, because of its high degree of redundancy, relatively short paths to it (see Feynman, 1965, p. 47)—and let it be redesigned by a Euclidean. In the name of elegance, the Euclidean will look for a small number of relatively powerful assumptions from which the rest may be derived. In so doing, he will eliminate redundant assumptions. The net effects will be twofold: (1) With a smaller number of assumptions taken as axioms, the mean number of steps in a derivation will increase, and can do so exponentially. This increased length of seriation will decrease reliability along any path to the conclusion. (2) Alternative or parallel ways of getting to a given conclusion will be substantially decreased as redundant assumptions are removed, and this decrease in "parallation" will also decrease the total reliability of the conclusion.

Each of these changes increases the unreliability of the structure, and both of them operating together produce a cumulative effect—if errors are possible, as I have supposed. Not only is the probability of failure of the structure greater after it has been Euclideanized, but the consequences of failure become more severe: with less redundancy, the failure of any given

component assumption is likely to inform a larger part of the structure. I will elaborate on this point shortly. It has not been studied before now (but see Glymour,1980) because of the dominance of the Cartesian Euclidean perspective and because of a key artifact of firstorder logic.

Formal models of theoretical structures characteristically start with the assumption that the structures contain no inconsistencies. As a normative ideal, this is fine; but as a description of real scientific theories, it is inadequate. Most or all scientific theories with which I am familiar contain paradoxes and inconsistencies, either between theoretical assumptions or between assumptions and data in some combination. (Usually these could be resolved if one knew which of several eminently plausible assumptions to give up, but each appears to have strong support; so the assumptions—and the inconsistencies—remain.) This feature of scientific theories has not until now (with the development of nonmonotonic logic) been modeled, because of the fear of total collapse. In firstorder logic, anything whatsoever follows from a contradiction; so systems which contain contradictions are regarded as useless.

But the total collapse suggested by firstorder logic (or by highly Euclidean structures with little redundancy) seems not to be a characteristic of scientific theories. The thing that is remarkable about scientific theories is that the inconsistencies are walled off and do not appear to affect the theory other than very locally— for things very close to and strongly dependent on one of the conflicting assumptions. Robustness provides a possible explanation, perhaps the best explanation, for this phenomenon.

When an inconsistency occurs, results which depend on one or more of the contradictory assumptions are infirmed. This infection is transitive; it passes to things that depend on these results, and to their logical descendants, like a string of dominoes—until we reach something that has independent support. The independent support of an assumption sustains it, and the collapse propagates no further. If all deductive or inferential paths leading from a contradiction pass through robust results, the collapse is bounded within them, and the inconsistencies are walled off from the rest of the network. For each robust result, one of its modes of support is destroyed; but it has others, and therefore the collapse goes no further. Whether this is the only mechanism by which this isolation of contradictions could be accomplished, I do not know, but it is a possible way, and scientific constructs do appear to have the requisite robustness. (I am not aware that anyone has tried to formalize or to simulate this, though Stuart A. Kauffman's work on "forcing structures" in binary, Boolean switching networks seems clearly relevant. See, for example, Kauffman, 1971, where these models are developed and applied to gene control networks. )

## Robustness, Testability, and the Nature of Theoretical Terms

Another area in which robustness is involved (and which is bound to see further development) is Clark Glymour's account of testing and evidential relations in theories. Glymour argues systematically that parts of a theoretical structure can be and are used to test other parts of the theory, and even themselves. (His name for this is bootstrapping.) This testing requires the determination of values for quantities of the theory in more than one way: "If the data are consistent with the theory, then these different computations must agree [within a tolerable experimental error] in the value they determine for the computed quantity; but if the data are inconsistent with the theory, then different computations of the same quantity may give different

results. Further and more important, what quantities in a theory may be computed from a given set of initial data depends both on the initial data and on the structure of the theory" (Glymour, 1980, p. 113).

Glymour argues later (pp. 139140) that the different salience of evidence to different hypotheses of the theory requires the use of a variety of types of evidence to test the different component hypotheses of the theory. Commenting on the possibility that one could fail to locate the hypothesis whose incorrectness is producing an erroneous determination of a quantity or, worse, mislocating the cause of the error, he claims: "The only means available for guarding against such errors is to have a variety of evidence so that as many hypotheses as possible are tested in as many different ways as possible. What makes one way of testing relevantly different from another is that the hypotheses used in one computation are different from the hypotheses used in the other computation. Part of what makes one piece of evidence relevantly different from another piece of evidence is that some test is possible from the first that is not possible from the second, or that, in the two cases, there is some difference in the precision of computed values of theoretical quantities" (Glymour, 1980, p. 140).

A given set of data and the structure of the theory permit a test of a hypothesis (or the conjunction of a group of hypotheses) if and only if they permit determination of all of the values in the tested entity in such a way that contradictory determinations of at least one of these values could result (in the sense that it is not analytically ruled out). This requires more than one way of getting at that value. (See Glymour, 1980, p. 307.) To put it in the language of the present paper, *only robust hypotheses are testable.* Furthermore, a theory in which most components are multiply connected is a theory whose faults are relatively precisely localizable. Not only do errors not propagate far, but we can find their source quickly and evaluate the damage and what is required for an adequate replacement. If this sounds like a design policy for an automobile, followed to facilitate easy diagnostic service and repair, I can say only that there is no reason why our scientific theories should be less well designed than our other artifacts.

The same issues arise in a different way in Campbell 's discussions (Campbell and Fiske, 1959; Campbell, 1969a, 1969b, 1977; Cook and Campbell, 1979) of single or definitional versus multiple operationalism. Definitional operationalism is the view that philosophers know as operationalism, that the meaning of theoretical terms is to be defined in terms of the experimental operations used in measuring that theoretical quantity. Multiple means of determining such a quantity represents a paradox for this view—an impossibility, since the means is definitive of the quantity, and multiple means means multiple quantities. Campbell 's multiple operationalism is not operationalism at all in this sense but a more tolerant and eclectic empiricism, for he sees the multiple operations as contingently associated with the thing measured. Being contingently associated, they cannot have a definitional relation to it; consequently, there is no barrier to accepting that one (robust) quantity has a number of different operations to get at it, each too imperfect to have a definitional role but together triangulating to give a more accurate and complete picture than would be possible from any one of them alone.

Campbell's attack on definitional operationalism springs naturally from his fallibilism and his critical realism. Both of these forbid a simple definitional connection between theoretical constructs and measurement operations: "One of the great weaknesses in definitional operationalism as a description of best scientific practice was that it allowed no formal way of expressing the scientist's prepotent awareness of the imperfection of his measuring instruments and his prototypic activity of improving them" (Campbell, 1969a, p. 15). For a realist the

connection between any measurement and the thing measured involves an often long and indirect causal chain, each link of which is affected and tuned by other theoretical parameters. The aim is to make the result insensitive to or to control these causally relevant but semantically irrelevant intermediate links: "What the scientist does in practice is to design the instrument so as to minimize and compensate for the stronger of these irrelevant forces. Thus, the galvanometer needle is as light as possible, to minimize inertia. It is set on jeweled bearings to minimize friction. It may be used in a leadshielded and degaussed room. Remote influences are neglected because they dissipate at the rate of l/d2, and the weak and strong nuclear forces dissipate even more rapidly. But these are practical minimizations, recognizable on theoretical grounds as incomplete" (1969a, pp. 1415).

The very same indirectness and fallibility of measurement that rule out definitional links make it advantageous to use multiple links: "[W] e have only *other invalid measures* against which to validate our tests; we have no 'criterion' to check them against.... A theory of the interaction of two theoretical parameters must be tested by imperfect exemplifications of each.... In this predicament, great inferential strength is added when each theoretical parameter is exemplified in 2 or more ways, each mode being as independent as possible of the other, as far as the theoretically irrelevant components are concerned. This general program can be designated *multiple operationalism"* (Campbell, 1969a, p. 15).

Against all this, then, suppose one did have only one means of access to a given quantity. Without another means of access, even if this means of access were not made definitional, statements about the value of that variable would not be independently testable. Effectively, they would be as if defined by that means of access. And since the variable was not connected to the theory in any other way, it would be an unobservable, a fifth wheel: anything it could do could be done more directly by its operational variable. It is, then, in Margenau's apt phrase, a peninsular concept (Margenau, 1950, p. 87), a bridge that leads to nowhere.

Philosophers often misleadingly lump this "peninsularity" and the existence of extra axioms permitting multiple derivations together as redundancy. The implication is that one should be equally disapproving of both. Presumably, the focus on errorfree systems leads philosophers to regard partially identical paths (the paths from a peninsular concept and from its "operational variable" to any consequence accessible from either) and alternative independent paths (robustness, bootstrapping, or triangulation) as equivalent—because they are seen as equally dispensable if one is dealing with a system in which errors are impossible. But if errors are possible, the latter kind of redundancy can increase the reliability of the conclusion; the former cannot.

A similar interest in concepts with multiple connections and a disdain for the trivially analytic, singly or poorly connected concept is to be found in Hilary Putnam's (1962) classic paper "The Analytic and the Synthetic." Because theoretical definitions are multiply connected law-cluster concepts, whose meaning is determined by this multiplicity of connections, Putnam rejects the view that such definitions are stipulative or analytic. Though for Putnam it is theoretical connections, rather than operational ones, which are important, he also emphasizes the importance of a multiplicity of them: "Lawcluster concepts are constituted not by a bundle of properties as are the typical general names [cluster concepts] like 'man' and 'crow,' but by a cluster of laws which, as it were, determine the identity of the concept. The concept 'energy' is an excellent sample.... It enters into a great many laws. It plays a great many roles, and these laws and inference roles constitute its meaning collectively, not individually. I want to suggest that

most of the terms in highly developed sciences are lawcluster concepts, and that one should always be suspicious of the claim that a principle whose subject term is a lawcluster concept is analytic. The reason that it is difficult to have an analytic relationship between lawcluster concepts is that . . . any one law can be abandoned without destroying the identity of the law-cluster concept involved" (p. 379).

Statements that are analytic are so for Putnam because they are singly connected, not multiply connected, and thus trivial: "Thus, it cannot 'hurt' if we decide always to preserve the law 'All bachelors are unmarried' . . . because bachelors are a kind of synthetic class. They are a 'natural kind' in Mill's sense. They are rather grouped together by ignoring all aspects except a single legal one. One is simply not going to find any . . . [other] laws about such a class" (p. 384).

Thus, the robustness of a concept or law—its multiple connectedness within a theoretical structure and (through experimental procedures) to observational results—has implications for a variety of issues connected with theory testing and change, with the reliability and stability of laws and the component parts of a theory, with the discovery and localization of error when they fail, the analyticsynthetic distinction, and accounts of the meaning of theoretical concepts. But these issues have focused on robustness in existing theoretical structures. It is also important in discovery and in the generation of new theoretical structures.

## Robustness, Redundancy, and Discovery

For the complex systems encountered in evolutionary biology and the social sciences, it is often unclear what is fundamental or trustworthy. One is faced with a wealth of partially conflicting, partially complementary models, regularities, constructs, and data sets with no clear set of priorities for which to trust and where to start. In this case particularly, processes of validation often shade into processes of discovery—since both involve a winnowing of the generalizable and the reliable from the special and artifactual. Here too robustness can be of use, as Richard Levins suggests in the passage which introduced me to the term:

> Even the most flexible models have artificial assumptions. There is always room for doubt as to whether a result depends on the essentials of a model or on the details of the simplifying assumptions. This problem does not arise in the more familiar models, such as the geographical map, where we all know that contiguity on the map implies contiguity in reality, relative distances on the map correspond to relative distances in reality, but color is arbitrary and a microscopic view of the map would only show the fibers of the paper on which it is printed. But in the mathematical models of population biology, it is not always obvious when we are using too high a magnification.
>
> Therefore, we attempt to treat the same problem with several alternative models, each with different simplifications, but with a common biological assumption. Then, if these models, despite their different assumptions, lead to similar results we have what we can call a robust theorem which is relatively free of the details of the model. Hence, our truth is the inter section of independent lies [Levins, 1966, p. 423].

Levins is here making heuristic use of the philosopher's criterion of logical truth as true in all possible worlds. He views robustness analysis as "sampling from a space of possible models" (1968, p. 7). Since one cannot be sure that the sampled models are representative of the space,

one gets no guarantee of logical truth but, rather, a heuristic (fallible but effective) tool for discovering empirical truths which are relatively free of the details of the various specific models.

Levins talks about the robustness of theorems or phenomena or consequences of the models rather than about the robustness of the models themselves. This is necessary, given his view that any single model makes a number of artifactual (and therefore nonrobust) assumptions. A theory would presumably be a conceptual structure in which many or most of the fundamental theorems or axioms are relatively robust, as is suggested by Levins' statement (1968, p. 7) "A theory is a cluster of models, together with their robust consequences."

If a result is robust over a range of parameter values in a given model or over a variety of models making different assumptions, this gives us some independence of knowledge of the exact structure and parameter values of the system under study: a prediction of this result will remain true under a variety of such conditions and parameter values This is particularly important in scientific areas where it may be difficult to determine the parameter values and conditions exactly.

*Robust theorems can thus provide a more trustworthy basis for generalization of the model or theory* and also, through their independence of many exact details, *a sounder basis for predictions from it.* Theory generalization is an important component of scientific change, and thus of scientific discovery.

Just as robustness is a guide for discovering trustworthy results and generalizations of theory, and distinguishing them from artifacts of particular models, it helps us to distinguish signal from noise in perception generally. Campbell has furnished us with many examples of the role of robustness and pattern matching in visual perception and its analogues, sonar and radar. In an early paper, he described how the pattern and the redundancy in a randomly pulsed radar signal bounced off Venus gave a new and more accurate measurement of the distance to that planet (Campbell, 1966).

The later visual satellite pictures of Mars and its satellite Deimos have provided an even more illuminating example, again described by Campbell (1977) in the unpublished William James Lectures (lecture 4, pp. 89 90). The now standard procedures of image enhancement involve combining a number of images, in which the noise, being random, averages out; but the signal, weak though usually present, adds in intensity until it stands out. The implicit principle is the same one represented explicitly in von Neumann's (1956) use of "majority organs" to filter out error: the combination of parallel or redundant signals with a threshold, in which it is assumed that the signal, being multiply represented, will usually exceed threshold and be counted; and the noise, being random, usually will fall below threshold and be lost. There is an art to designing the redundancy so as to pick up the signal and to setting the threshold so as to lose the noise. It helps, of course, if one knows what he is looking for. In this case of the television camera centered on Mars, Deimos was a moving target and—never being twice in the same place to add appropriately (as were the static features of Mars)—was consequently filtered out as noise. But since the scientists involved knew that Deimos was there, they were able to fix the image enhancement program to find it. By changing the threshold (so that Deimos and some noise enter as— probably smeared—signal), changing the sampling rate or the integration area (stopping Deimos at the effectively same place for two or more times), or introducing the right kind of spatiotemporal correlation function (to track Deimos's periodic moves around Mars), the~~~ could restore Deimos to the pictures again. Different tunings of the noise filters and

different redundancies in the signal were exploited to bring static Mars and moving Deimos into clear focus.

We can see exactly analogous phenomena in vision if we look at a moving fan or airplane propeller. We can look through it (filtering it out as noise) to see something behind it. Lowering our threshold, we can attend to the propeller disk as a colored transparent (smeared) object. Crossspecific variation in flickerfusion frequency indicates different sampling rates, which are keyed to the adaptive requirements of the organism (see Wimsatt, 1980a, pp. 292297). The various phenomena associated with periodic stroboscopic illumination (apparent freezing and slow rotation of a rapidly spinning object) involve detection of a lagged correlation. Here, too, different tunings pick out different aspects of or entities in the environment. This involves a use of different heuristics, a matter I will return to later.

I quoted Glymour earlier on the importance of getting the same answer for the value of quantities computed in two different ways. What if these computations or determinations do not agree? The result is not always disastrous; indeed, when such mismatches happen in a sufficiently structured situation, they can be very productive.

This situation could show that we were wrong in assuming that we were detecting or determining the same quantity; but (as Campbell, 1966, was the first to point out), if we assume that we *are* determining the same quantity but "through a glass darkly," the mismatch can provide an almost magical opportunity for discovery. Given imperfect observations of a thingweknownotwhat, using experimental apparatus with biaseswemaynotunderstand, we can achieve both a better understanding of the object (it must be, after all, that one thing whose properties can produce these divergent results in these detectors) and of the experimental apparatus (which are, after all, these pieces that can be affected thus divergently by this one thing).

The constraint producing the information here is the identification of the object of the two or more detectors. If two putatively identical things are indeed identical, then any property of one must be a property of the other. We must resolve any apparent differences either by giving up the identification or locating the differences not in the thing itself but in the interactions of the thing with different measuring instruments. And this is where we learn about the measuring instruments. Having then acquired a better knowledge of the biases of the measuring instruments, we are in a better position not only to explain the differences but also, in the light of them, to give a newly refined estimate of the property of the thing itself. This procedure, a kind of "means-end" analysis (Wimsatt, 1976a; Simon, 1969) has enough structure to work in any given case only because of the enormous amount of background knowledge of the thing and the instruments which we bring to the situation. What we can learn (in terms of localizing the source of the differences) is in direct proportion to what we already know.

This general strategy for using identifications has an important subcase in reductive explanation. I have argued extensively (Wimsatt, 1976a, part II, 1976b, 1979) that the main reason for the productiveness of reductive explanation is that interlevel identifications immediately provide a wealth of new hypotheses: each property of the entity as known at the lower level must be a property of it as known at the upper level, and conversely; and usually very few of these properties from the other level have been predicated of the common object. The implications of these predictions usually have fertile consequences at both levels, and even where the match is not exact, there is often enough structure in the situation to point to a revised identification, with the needed refinements. This description characterizes well the history of

genetics, both in the period of the localization of the genes on chromosomes (1883 to 1920) and in the final identification of DNA as the genetic material (1927 to 1953). (For the earlier period see, for example, Allen, 1979; Moore, 1972; Darden, 1974; Wimsatt, 1976a, part II. For the later period see Olby, 1974.) Indeed, the overall effect of these considerations is to suggest that *the use of identities for the deletion of error in a structured situation for the detection of error may be the most powerful heuristic known and certainly one of the most effective in generating scientific hypotheses.*

Also significant in the connection between robustness and discovery is Campbell's (1977) suggestion that things with greater entitativity (things whose boundaries are more robust) ought to be learned earlier. He cites suggestive support from language development for this thesis, which Quine's (1960) views also tend to support. I suspect that robustness could prove to be an important tool in analyzing not only what is discovered but also the order in which things are discovered.

There is some evidence from work with children (Omanson, 1980a, 1980b) that components of narratives which are central to the narrative, in that they are integrated into its causal and its purposive or intentional structure, are most likely to be remembered and least likely to be abstracted out in summaries of the story. This observation is suggestively related both to Feynman's (1965, p. 47) remark quoted above, relating robustness to forgetting relationships in a multiply connected theory, and to Simon's (1969) concept of a blackboard work space, which is maintained between successive attempts to solve a problem and in which the structure of the problem representation and goal tree may be subtly changed through differential forgetting. These suggest other ways in which robustness could affect discovery processes through differential effects on learning and forgetting.

## Robustness, Objectification, and Realism

Robustness is widely used as a criterion for the reality or trustworthiness of the thing which is said to be robust. The boundaries of an ordinary object, such as a table, as detected in different sensory modalities (visually, tactually, aurally, orally), roughly coincide, making them robust; and this is ultimately the primary reason why we regard perception of the object as veridical rather than illusory. (See Campbell, 1958, 1966.) It is a rare illusion indeed which could systematically affect all of our senses in this consistent manner. (Drug induced hallucinations and dreams may involve multimodal experience but fail to be consistent through time for a given subject, or across observers, thus failing at a higher level to show appropriate robustness.)

Our concept of an object is of something which exemplifies a multiplicity of properties within its boundaries, many of which change as we move across its boundary. A onedimensional object is a contradiction in terms and usually turns out to be a disguised definition—a legal or theoretical fiction. In appealing to the robustness of boundaries as a criterion for objecthood, we are appealing to this multiplicity of properties (different properties detected in different ways) and thus to a timehonored philosophical notion of objecthood.

Campbell (1958) has proposed the use of the coincidence of boundaries under different means of detection as a methodological criterion for recognizing entities such as social groups. For example, in a study of factors affecting the reproductive cycles of women in college dormitories, McClintock (1971, and in conversation) found that the initially randomly timed and differentlength cycles of 135 women after several months became synchronized into 17 groups,

each oscillating synchronously, in phase and with a common period. The members of these groups turned out to be those who spent most time together, as determined by sociological methods. After the onset of synchrony, group membership of an individual could be determined either from information about her reproductive cycle or from a sociogram representing her frequency of social interaction with other individuals. These groups are thus multiply detectable. This illustrates the point that there is nothing sacred about using perceptual criteria in individuating entities. The products of any scientific detection procedure, including procedures drawn from different sciences, can do as well, as Campbell suggests: "In the diagnosis of middlesized physical entities, the boundaries of the entity are multiply confirmed, with many if not all of the diagnostic procedures confirming each other. For the more 'real' entities, the number of possible ways of confirming the boundaries is probably unlimited, and the more our knowledge expands, the more diagnostic means we have available. 'Illusions' occur when confirmation is attempted and found lacking, when boundaries diagnosed by one means fail to show up by other expected checks" (1958, pp. 2324).

Illusions can arise in connection with robustness in a variety of ways. Campbell's remark points to one important way: Where expectations are derived from one boundary, or even more, the coincidence of several boundaries leads us to predict, assume, or expect that other relevant individuating boundaries will coincide. Perhaps most common, given the reductionism common today, are situations in which the relevant system boundary is in fact far more inclusive than one is led to expect from the coincidence of a number of boundaries individuating an object at a lower level. Such functional localization fallacies are found in neurophysiology, in genetics, in evolutionary biology (with the hegemony of the selfish gene at the expense of the individual or the group; see Wimsatt, 1980b), in psychology, and (where it is a fallacy) with methodological individualism in the social sciences. In all these cases the primary object of analysis—be it a gene, a neuron, a neural tract, or an individual—may well be robust, but its high degree of entitativity leads us to hang too many boundaries and explanations on it. Where this focal entity is at a lower level, reductionism and robustness conspire to lead us to regard the higherlevel systems as epiphenomenal. Another kind of illusion—the illusion that an entity is robust—can occur when the various means of detection supposed to be independent are not in fact. (This will be discussed further in the final section of this chapter.) Another kind of illusion or paradox arises particularly for functionally organized systems. This illusion occurs when a system has robust boundaries, but the different criteria used to decompose it into parts produce radically different boundaries. When the parts have little entitativity compared to the system, the holist's war cry (that the whole is more than the sum of the parts) will have a greater appeal. Elsewhere (Wimsatt, 1974), I have explored this kind of case and its consequences for the temptation of antireductionism, holism, or, in extreme cases, vitalisms or ontological dualisms.

Robustness is a criterion for the reality of entities, but it also has played and can play an important role in the analysis of properties. Interestingly, the distinction between primary and secondary qualities, which had a central role in the philosophy of Galileo, Descartes, and Locke, can be made in terms of robustness. Primary qualities—such as shape, figure, and size—are detectable in more than one sensory modality. Secondary qualities—such as color, taste, and sound—are detectable through only one sense. I think it is no accident that seventeenthcentury philosophers chose to regard primary qualities as the only things that were "out there"— in objects; their crossmodal detectability seemed to rule out their being products of sensory interaction with the world. By contrast the limitation of the secondary qualities to a single

sensory modality seemed naturally to suggest that they were "in us," or subjective. Whatever the merits of the further seventeenth-century view that the secondary qualities were to be explained in terms of the interaction of a perceiver with a world of objects with primary qualities, this explanation represents an instance of an explanatory principle which is widely found in science (though seldom if ever explicitly recognized): *the explanation of that which is not robust in terms of that which is robust*. (For other examples see Wimsatt, 1976a, pp. 243249; Feynman, 1965).

Paralleling the way in which Levins' use of robustness differs from Feynman's, *robustness, or the lack of it, has also been used in contexts where we are unsure about the status of purported properties, to argue for their veridicality or artifactuality,* and thus to discover the properties in terms of which we should construct our theories. This is the proposal of the now classic and widely used methodological paper of Campbell and Fiske (1959). Their convergent validity is a form of robustness, and their criterion of discriminant validity can be regarded as an attempt to guarantee that the invariance across test methods and traits is not due to their insensitivity to the variables under study. Thus, method bias, a common cause of failures of discriminant validity, is a kind of failure of the requirement for robustness that the different means of detection used are actually independent, in this case because the method they share is the origin of the correlations among traits.

Campbell and Fiske point out that very few theoretical constructs (proposed theoretical properties or entities) in the social sciences have significant degrees of convergent and discriminant validity, and they argue that this is a major difference between the social and natural or biological sciences—a difference which generates many of the problems of the social sciences. (For a series of essays which in effect claim that personality variables are highly context dependent and thus have very little or no robustness, see Shweder, 1979a, 1979b, 1980.)

While the natural and biological sciences have many problems where similar complaints could be made (the importance of interaction effects and context dependence is a key indicator of such problems), scientists in these areas have been fortunate in having at least a large number of cases where the systems, objects, and properties they study can be effectively isolated and localized, so that interactions and contexts can be ignored.

## Robustness and Levels of Organization

Because of their multiplicity of connections and applicable descriptions, robust properties or entities tend to be (1) more easily detectable, (2) less subject to illusion or artifact, (3) more ex*planatorily fruitful, and (4) predictively richer than nonrobust properties or entities.* With this set of properties, it should be small wonder that we use robustness as a criterion for reality. It should also not be surprising that—since we view perception (as evolutionary epistemologists do) as an efficient tool for gathering information about the world—robustness should figure centrally in our analysis of perceptual hypotheses and heuristics (in the earlier section "Robustness, Redundancy, and Discovery" and in the next section, "Heuristics and Robustness"). Finally, since ready detectability, relative insensitivity to illusion or artifact, and explanatory and predictive fruitfulness are desirable properties for the components of scientific theories, we should not be surprised to discover that robustness is important in the discovery and description of phenomena (again, see the section on discovery) and in analyzing the structure of scientific theories (see the section "Robustness and the Structure of Theories").

One of the most ubiquitous phenomena of nature is its tendency to come in levels. If the aim of science, to follow Plato, is to cut up nature at its joints, then these levels of organization must be its major vertebrae. They have become so major, indeed, that our theories tend to follow these levels, and the language of our theories comes in strata. This has led many linguistically inclined philosophers to forgo talk of nature at all, and to formulate problems—for example, problems of reduction—in terms of "analyzing the relation between theoretical vocabularies at different levels." But our language, as Campbell (1974) would argue, is just another (albeit very important) tool in our struggle to analyze and to adapt to nature. In an earlier paper (Wimsatt, 1976a, part III), I applied Campbell's criteria for entification to argue that entities at different levels of organization tend to be multiply connected in terms of their causal relations, primarily with other entities at their own level, and that they, and the levels they comprise, are highly robust. As a result, there are good explanatory reasons for treating different levels of organization as dynamically, ontologically, and epistemologically autonomous. There is no conflict here with the aims of good reductionistic science: there is a great deal to be learned about upperlevel phenomena at lower levels of organization, but upperlevel entities are not "analyzed away" in the process, because they remain robustly connected with other upper level entities, and their behavior is explained by upperlevel variables.

To see how this is so, we need another concept—that of the *sufficient parameter,* introduced by Levins (1966, pp. 428, 429):

> It is an essential ingredient in the concept of levels of phenomena that there exists a set of what, by analogy with the sufficient statistic, we can call sufficient parameters defined on a given level... which are very much fewer than the number of parameters on the lower level and which among them contain most of the important information about events on that level.
>
> The sufficient parameters may arise from the combination of results of more limited studies. In our robust theorem on niche breadth we found that temporal variation, patchiness of the environment, productivity of the habitat, and mode of hunting could all have similar effects and that they did this by way of their contribution to the uncertainty of the environment. Thus uncertainty emerges as a sufficient parameter.
>
> The sufficient parameter is a manytoone transformation of lowerlevel phenomena. Therein lies its power and utility, but also a new source of imprecision. The manytoone nature of "uncertainty" prevents us from going backwards. If either temporal variation or patchiness or low productivity leads to uncertainty, the consequences of uncertainty alone cannot tell us whether the environment is variable, or patchy, or unproductive. Therefore, we have lost information.

A sufficient parameter is thus a parameter, a variable, or an index which, either for most purposes or merely for the purposes at hand, captures the effect of significant variations in lowerlevel or less abstract variables (usually only for certain ranges of the values of these variables) and can thus he substituted for them in the attempt to build simpler models of the upperlevel phenomena.

Levins claims that this notion is a natural consequence of the concept of levels of phenomena, and this is so, though it may relate to degree of abstraction as well as to degree of aggregation. (The argument I will give here applies only to levels generated by aggregation of lowerlevel entities to form upperlevel ones.) Upper-level variables, which give a more

"coarsegrained" description of the system, are much smaller in number than the lowerlevel variables necessary to describe the same system. Thus, there must be, for any given degree of resolution between distinguishable state descriptions, far fewer distinguishable upperlevel state descriptions than lowerlevel ones. The smaller number of distinguishable upperlevel states entails that for any given degree of resolution, there must be manyone mappings between at least some lower-level and upperlevel state descriptions with many lowerlevel descriptions corresponding to a single upperlevel description. But then, those upperlevel state descriptions with multiple lowerlevel state descriptions are robust over changes from one of these lower-level descriptions to another in its set.

Furthermore, the stability of (and possibility of continuous change in) upperlevel phenomena (remaining in the same macrostate or changing by moving to neighboring states) places constraints on the possible mappings between lowerlevel and upper-level states: in the vast majority of cases neighboring microstates must map without discontinuity into the same or neighboring macrostates; and, indeed, most local microstate changes will have no detectable macrolevel effects. *This fact gives upperlevel phenomena and laws a certain insulation from* (through their invariance over: robustness again!) *lower-level changes and generates a kind of explanatory and dynamic (causal) autonomy of the upper-level phenomena and processes,* which I have argued for elsewhere (Wimsatt, 1976a, pp. 249251; 1976b).

If one takes the view that causation is to be characterized in terms of manipulability (see, for example, Gasking, 1955; Cook and Campbell, 1979), the fact that the vast majority of manipulations at the microlevel do not make a difference at the macrolevel means that macrolevel variables are almost always more causally efficacious in making macrolevel changes than microlevel variables. This gives explanatory and dynamic autonomy of the upperlevel entities, phenomena, laws, and relations, within a view of explanation which is sensitive to problems of computational complexity and the costs and benefits we face in offering explanations. As a result, it comes much closer than the traditional hypotheticodeductive view to being able to account for whether we explain a phenomenon at one level and when we choose to go instead to a higher or lower level for its explanation. (See Wimsatt, 1976a, part III, and 1976b, particularly sections 4, 5, 6, and the appendix.)

The manyone mappings between lower and upperlevel state descriptions mentioned above are consistent with correspondences between types of entities at lower and upper levels but do not entail them. There may be only tokentoken mappings (piece-meal mappings between instances of concepts, without any general mappings between concepts), resulting in the upperlevel proper ties being supervenient on rather than reducible to lowerlevel properties (Kim, 1978; Rosenberg, 1978). The main difference between Levins' notion of a sufficient parameter and the notion of supervenience is that the characterization of supervenience is embedded in an assumed apocalyptically complete and correct description of the lower and upper levels. Levins makes no such assumption and defines the sufficient parameter in terms of the imperfect and incomplete knowledge that we actually have of the systems we study. It is a broader and less demanding notion, involving a relation which is inexact, approximate, and admits of both unsystematic exceptions (requiring a *ceteris paribus* qualifier) and systematic ones (which render the relationship conditional).

Supervenience could be important for an omniscient Laplacean demon but not for real, fallible, and limited scientists. The notion of supervenience could be regarded as a kind of ideal limiting case of a sufficient parameter as we come to know more and more about the system, but

it is one which is seldom if ever found in the models of science. The concept of a sufficient parameter, by contrast, has many instances in science. It is central to the analysis of reductive explanation (Wimsatt, 1976a; 1976b, pp. 685689; 1979) and has other uses as well (Wimsatt, 1980a, section 4).

## Heuristics and Robustness

Much or even most of the work in philosophy of science today which is not closely tied to specific historical or current scientific case studies embodies a metaphysical stance which, in effect, assumes that the scientist is an omniscient and computationally omnipotent Laplacean demon. Thus, for example, discussions of reductionism are full of talk of "in principle analyzability" or "in principle deducibility," where the force of the "in principle" claim is held to be something like "If we knew a total description of the system at the lower level, and all the lowerlevel laws, a sufficiently complex computer could generate the analysis of all the upperlevel terms and laws and predict any upperlevel phenomenon." Parallel kinds of assumptions of omniscience and computational omnipotence are found in rational decision theory, discussions of Bayesian epistemology, automata theory and algorithmic procedures in linguistics and the philosophy of mind, and the reductionist and foundationalist views of virtually all the major figures of twentiethcentury logical empiricism. It seems almost to be a corollary to a deductivist approach to problems in philosophy of science (see Wimsatt, 1979) and probably derives ultimately from the Cartesian vision criticized earlier in this chapter.

I have already written at some length attacking this view and its application to the problem of reduction in science (see Wimsatt, 1974; 1976a, pp. 219237; 1976b; 1979; 1980b, section 3; and also Boyd,1972). The gist of this attack is threefold: (1) On the "Laplacean demon" interpretation of "in principle" claims, we have no way of evaluating their warrant, at least in science. (This is to be distinguished from cases in mathematics or automata theory, where "in principle" claims can be explicated in terms of the notion of an effective procedure.) (2) We are in any case not Laplacean demons, and a philosophy of science which could have normative force only for Laplacean demons thus gives those of us who do not meet these demanding specifications only counterfactual guidance; that is, it is of no real use to practicing scientists and, more strongly, suggests methods and viewpoints which are less advantageous than those derived from a more realistic view of the scientist as problem solver (see Wimsatt, 1979). (3) An alternative approach, which assumes more modest capacities of practicing scientists, does provide real guidance, better fits with actual scientific practice, and even (for reductive explanations) provides a plausible an d attractive alternative interpretation for the "in principle" talk which so many philosophers and scientists use frequently (see Wimsatt, 1976a, part II; 1976b, pp. 697701).

An essential and pervasive feature of this more modest alternative view is the replacement of the vision of an ideal scientist as a computationally omnipotent algorithmizer with one in which the scientist as decision maker, while still highly idealized, must consider the size of computations and the cost of data collection, and in other very general ways must be subject to considerations of efficiency, practical efficacy, and costbenefit constraints. This picture has been elaborated over the last twentyfive years by Herbert Simon and his coworkers, and their ideal is "satisficing man," whose rationality is bounded, by contrast with the unbounded omniscience and computational omnipotence of the "economic man" of rational decision theory (see Simon, 1957,

reprinted as ch. 1 of Simon, 1979; see also Simon, 1969). Campbell's brand of fallibilism and critical realism from an evolutionary perspective also place him squarely in this tradition.

A key feature of this picture of man as a boundedly rational decision maker is the use of heuristic principles where no algorithms exist or where the algorithms that do exist require an excessive amount of information, computational power, or time. I take a heuristic procedure to have three important properties (see also Wimsatt, 1980b, section 3): (1) By contrast with an algorithmic procedure (here ignoring probabilistic automata), *the correct application of a heuristic procedure does not guarantee a solution* and, if it produces a solution, does not guarantee that the solution is correct. (2) *The expected time, effort, and computational complexity of producing a solution with a heuristic procedure is appreciably less* (often by many orders of magnitude for a complex problem) *than that expected with an algorithmic procedure.* This is indeed the reason why heuristics are used. They are a cost-effective way, and often the *only* physically possible way, of producing a solution. (3) *The failures and errors produced when a heuristic is used are not random but systematic. I* conjecture that *any heuristic, once we understand how it works, can be made to fail.* That is, given this knowledge of the heuristic procedure, we can construct classes of problems for which it will always fail to produce an answer or for which it will always produce the wrong answer. This property of systematic production of wrong answers will be called the *bias* of the heuristic.

This last feature is exceedingly important. Not only can we work forward from an understanding of a heuristic to predict its biases, but we can also work backward from the observation of systematic biases as data to hypothesize the heuristics which produced them; and if we can get independent evidence (for example, from cognitive psychology) concerning the nature of the heuristics, we can propose a wellfounded explanatory and predictive theory of the structure of our reasoning in these areas. This approach was implicitly (and sometimes explicitly) followed by Tversky and Kahneman (1974), in their analysis of fallacies of probabilistic reasoning and of the heuristics which generate them (see also Shweder, 1977, 1979a, 1979b, 1980, for further applications of their work; and Mynatt, Doherty, and Tweney, 1977, for a further provocative study of bias in scientific reasoning). The systematic character of these biases also allows for the possibility of modifications in the heuristic or in its use to correct for them (see Wimsatt, 1980b, pp. 5254).

The notion of a heuristic has far greater implications than can be explored in this chapter. In addition to its centrality in human problem solving, it is a pivotal concept in evolutionary biology and in evolutionary epistemology. It is a central concept in evolutionary biology because any biological adaptation meets the conditions given for a heuristic procedure. First, it is a commonplace among evolutionary biologists that adaptations, even when functioning properly, do not guarantee survival and production of offspring. Second, they are, however, costeffective ways of contributing to this end. Finally, any adaptation has systematically specifiable conditions, derivable through an understanding of the adaptation, under which its employment will actually decrease the fitness of the organism employing it, by causing the organism to do what is, under those conditions, the wrong thing for its survival and reproduction. (This, of course, seldom happens in the organism's normal environment, or the adaptation would become maladaptive and be selected against.) This fact is indeed systematically exploited in the functional analysis of organic adaptations. It is a truism of functional inference that learning the conditions under which a system malfunctions, and how it malfunctions under those conditions, is a powerful tool for determining how it functions normally and the conditions under which it

was designed to function. (For illuminating discussions of the problems, techniques, and fallacies of functional inference under a variety of circumstances, see Gregory, 1962;Lorenz, 1965; Valenstein, 1973; Glassman, 1978.)

The notion of a heuristic is central to evolutionary epistemology because Campbell's (1974, 1977) notion of a vicarious selector, which is basic to his conception of a hierarchy of adaptive and selective processes spanning subcognitive, cognitive, and social levels, is that of a heuristic procedure. For Campbell a vicarious selector is a substitute and less costly selection procedure acting to optimize some index which is only contingently connected with the index optimized by the selection process it is substituting for. This contingent connection allows for the possibility—indeed, the inevitability—of systematic error when the conditions for the contingent concilience of the substitute and primary indices are not met. An important ramification of Campbell's idea of a vicarious selector is the possibility that one heuristic may substitute for another (rather than for an algorithmic procedure) underrestricted sets of conditions, and that this process may be repeated, producing a nested hierarchy of heuristics. He makes ample use of this hierarchy in analyzing our knowing processes (Campbell, 1974). I believe that this is an appropriate model for describing the nested or sequential structure of many approximation techniques, limiting operations, and the families of progressively mole realistic models found widely in progressive research programs, as exemplified in the development of nineteenthcentury kinetic theory, early twentiethcentury genetics, and several areas of modern population genetics and evolutionary ecology.

To my mind, Simon's work and that of Tversky and Kahneman have opened up a whole new set of questions and areas of investigation of pragmatic inference (and its informal fallacies) in science, which could revolutionize our discipline in the next decade. (For a partial view of how studies of reduction and reductionism in science could be changed, see Wimsatt, 1979.) This change in perspective would bring philosophy of science much closer to actual scientific practice without surrendering a normative role to an allembracing descriptivism. And it would reestablish ties with psychology through the study of the character, limits, and biases of processes of empirical reasoning. Inductive procedures in science are heuristics (Shimony, 1970), as are Mill's methods and other methods for discovering causal relations, building models, and generating and modifying hypotheses.

Heuristics are also important in the present context, because the procedures for determining robustness and for making further application of these determinations for other ends are all heuristic procedures. Robustness analysis covers a class of powerful and important techniques, but they are not immune to failures. There are no magic bullets in science, and these are no exception.

Most striking of the ways of failure of robustness analysis is one which produces illusions of robustness: the failure of the different supposedly independent tests, means of detection, models, or derivations to be truly independent. This is the basis for a powerful criticism of the validity of IQ scales as significant measures of intelligence (see McClelland, 1973). Failures of independence are not easy to detect and often require substantial further analysis. Without that, such failures can go undetected by the best investigators for substantial lengths of time. Finally, the fact that different heuristics can be mutually reinforcing, each helping to hide the biases of the others (see Wimsatt, 1980b, sections 5 and 8), can make it much harder to detect errors which would otherwise lead to discovery of failures of independence. The failure of independence in its various modes, and the factors affecting its discovery, emerges as one of the most critical and

important problems in the study of robustness analysis, as is indicated by the history of the group selection controversy.

## Robustness, Independence, and Pseudorobustness: A Case Study

In recent evolutionary biology (since Williams' seminal work in 1966), group selection has been the subject of widespread attack and general suspicion. Most of the major theorists (including W. D. Hamilton, John Maynard Smith, and E. O. Wilson) have argued against its efficacy. A number of mathematical models of this phenomenon have been constructed, and virtually all of them (see Wade, 1978) seem to support this skepticism. The various mathematical models of group selection surveyed by Wade all admit of the possibility of group selection. But almost all of them predict that group selection should only very rarely be a significant evolutionary factor; that is, they predict that group selection should have significant effects only under very special circumstances—for extreme values of parameters of the models—which should seldom be found in nature. Wade undertook an experimental test of the relative efficacy of individual and group selection—acting in concert or in opposition in laboratory populations of the flour beetle, *Tribolium.* This work produced surprising results. Group selection appeared to be a significant force in these experiments, one capable of overwhelming individual selection in the opposite direction for a wide range of parameter values. This finding, apparently contradicting the results of all of the then extant mathematical models of group selection, led Wade (1978) to a closer analysis of these models, with results described here.

All the models surveyed made simplifying assumptions, most of them different. Five assumptions, however, were widely held in common; of the twelve models surveyed, each made at least three of these assumptions, and five of the models made all five assumptions. Crucially, for present purposes, the five assumptions are biologically unrealistic and incorrect, and each independently has a strong negative effect on the possibility or efficacy of group selection. It is important to note that these models were advanced by a variety of different biologists, some sympathetic to and some skeptical of group selection as a significant evolutionary force. Why, then, did all of them make assumptions strongly inimical to it? Such a coincidence, radically improbable at best, cries out for explanation: we have found a systematic bias suggesting the use of a heuristic.

These assumptions are analyzed more fully elsewhere (Wade, 1978; Wimsatt, 1980a). My discussion here merely summarizes the results of my earlier analysis, where (in section 5) I presented a list of nine reductionistic research and modeling strategies. Each is a heuristic in that it has systematic biases associated with it, and these biases will lead to the wrong answer if the heuristic is used to analyze certain kinds of systems. It is the use of these heuristics, together with certain "perceptual" biases (deriving from thinking of groups as "collections of individuals" rather than as robust entities analogous to organisms), that is responsible for the widespread acceptance of these assumptions and the almost total failure to notice what an unrealistic view they give of group selection. Most of the reductionistic heuristics lead to a dangerous oversimplification of the environment being studied and a dangerous underassessment of the effects of these simplifications. In the context of the perceptual bias of regarding groups as collections of individuals (or sometimes even of genes), the models tend systematically to err in

the internal and relational structure they posit for the groups and in the character of processes of group reproduction and selection.

The first assumption, that the processes can be analyzed in terms of selection of alternative alleles at a single locus, is shown to be empirically false by Wade's own experiments, which show conclusively that both individual and group selection is proceeding on multilocus traits. (For an analysis of the consequences of treating a multilocus trait erroneously as a singlelocus trait, see Wimsatt, 1980b, section 4.) The fifth assumption, that individual and group selections are opposed in their effects, also becomes untenable for a multilocus trait (see Wimsatt, 1980b, section 7).

The second assumption is equivalent to the timehonored assumption of panmixia, or random mating within a population, but in the context of a group selection model it is equivalent to assuming a particularly strong form of blending inheritance for group inheritance processes. This assumption is factually incorrect and, as R. A. Fisher showed in 1930, effectively renders evolution at that level impossible. The third assumption is equivalent to assuming that groups differ in their longevity but not in their reproductive rates. But, as all evolutionary biologists since Darwin have been aware, variance in reproductive rate has a far greater effect on the intensity of selection than variance in longevity. So the more significant component was left out in favor of modeling the less significant one. (The second and third assumptions are discussed in Wimsatt, 1980b, section 7.) The fourth assumption is further discussed and shown to be incorrect in Wade (1978).

The net effect is a set of cumulatively biased and incorrect assumptions, which, not surprisingly, lead to the incorrect conclusion that group selection is not a significant evolutionary force. If I am correct in arguing that these assumptions probably went unnoticed because of the biases of our reductionistic research heuristics, a striking analogy emerges. The phenomenon appeared to be a paradigmatic example of Levinsian robustness. A wide variety of different models, making different assumptions, appeared to show that group selection could not be efficacious. But the robustness was illusory, because the models were not independent in their assumptions. The commonality of these assumptions appears to be a species of method bias, resulting in a failure of discriminant validity (Campbell and Fiske, 1959). But the method under consideration is not the normal sort of test instrument that social scientists deal with. Instances of the method are reductionistic research heuristics, and the method is reductionism. For the purposes of problem solving, our minds can be seen as a collection of methods, and the particularly singleminded are unusually prone to method bias in their thought processes. This conclusion is ultimately just another confirmation at another level of something Campbell has been trying to teach us for years about the importance of multiple independent perspectives.

## References

Allen, G. E. Thomas Hunt Morgan: The Man and His Science. Princeton, NJ.: Princeton University Press, 1979.

Barlow, R. E., and Proschan, F. The Mathematical Theory of Reliability and Life Testing New York: Wiley, 1975.

Boyd, R. "Determinism, Laws, and Predictability in Principle." Philosophy of Science, 1972, 39, 431450.

Campbell, D. T. "Common Fate, Similarity, and Other Indices of the Status of Aggregates of Persons as Social Entities." Behavioral Science, 1958,3, 1425.
Campbell, D. T. "Pattern Matching as an Essential in Distal Knowing." In K. R. Hammond (Ed.), The Psychology of Egon Brunswik. New York: Holt, Rinehart and Winston, 1966.
Campbell, D. T. "Definitional Versus Multiple Operationalism." et al., 1969a, 2 (1), 1417.
Campbell, D. T. "Prospective: Artifact and Control." In R. Rosenthal and R. Rosnow (Eds.), Artifact in Behavioral Research. New York: Academic Press, 1969b.
Campbell, D. T. "Evolutionary Epistemology." In P. A. Schilpp (Ed.), The Philosophy of Karl Popper. La Salle, lll: Open Court, 1974.
Campbell, D. T. "Descriptive Epistemology: Psychological, Sociological, and Evolutionary." William James Lectures, Harvard University, 1977.
Campbell, D. T., and Fiske, D. W. "Convergent and Discriminant Validation by the MultitraitMultimethod Matrix." Psychological Bulletin, 1959, 56, 81 105.
Cook, T. D., and Campbell, D. T. QuasiExperimentation: Design and Analysis for Field Settings. Chicago: Rand McNally, 1979.
Cronbach, L. J., and Meehl, P. E. "Construct Validity in Psychological Tests." Psychological Bulletin, 1955, 52, 281302.
Darden, L. "Reasoning in Scientific Change: The Field of Genetics at Its Beginnings." Unpublished doctoral dissertation, Committee on the Conceptual Foundations of Science, University of Chicago, 1974.
Feynman, R. P. The Character of Physical Law. Cambridge, Mass.: M.I.T. Press, 1965.
Fisher, R. A. The Genetically Theory of Natural Selection. New York: Clarendon Press, 1930.
Gasking, D. A. T. "Causation and Recipes." Mind, 1955, n.s. 64, 479487.
Glassman, R. B. "The Logic of the Lesion Experiment and Its Role in the Neural Sciences." In S. Finger (Ed.), Recovery from Brain Damage: Research and Theory. New York: Plenum, 1978.
Glymour, C. Theory and Evidence. Princeton, N.J.: Princeton University Press, 1980.
Gregory, R. L. "Models and the Localization of Function in the Central Nervous System" [1962]. Reprinted in C. R. Evans and A. D. J. Robertson (Eds.), Key Papers: Cybernetics. London: Butterworth, 1967.
Kauffman, S. A. "Cellular Gene Control Systems." In A. A. Moscona and others (Eds.), Current Topics in Developmental Biology. Vol. 6. New York: Academic Press, 1971.
Kim, J. "Supervenience and Nomological Incommensurables. " American Philosophical Quarterly, 1978, 15, 149156.
Laudan, L. "William Whewell on the Consilience of Inductions." The Monist, 1971, 55, 368391.
Levins, R. "The Strategy of Model Building in Population Biology." American Scientist, 1966, 54, 421 431.
Levins, R. Euolution in Changing Environments. Princeton, NJ.: Princeton University Press, 1968.

Lorenz, K. Z. Evolution and Modification of Behavior. Chicago: University of Chicago Press, 1965.
McClelland, D. D. "Testing for Competence Rather Than for 'Intelligence.' " American Psychologist, 1973, 29, 107.
McClintock, M. K. "Menstrual Synchrony and Suppression." Nature, January 22, 1971, 229, 244245.
Margenau, H. The Nature of Physical Reality. New York: McGrawHill, 1950.

Moore, J. A. Heredity and Development. (2nd ed.) New York: Oxford University Press, 1972.
Mynatt, C. R., Doherty, M. E., and Tweney, R. D. "Confirmation Bias in a Simulated Research Environment: An Experimental Study of Scientific Inference." Quarterly Journal of Experimental Psychology, 1977, 29, 8595.
Olby, R. The Path to the Double Helix. Seattle: University of Washington Press, 1974.
Omanson, R. C. "The Narrative Analysis: Identifying Central, Supportive and Distracting Content." Unpublished manuscript, 1980a.
Omanson, R. C. "The Effects of Centrality on Story Category Saliency: Evidence for Dual Processing." Paper presented at 88th annual meeting of the American Psychological Association, . Montreal, September 1980b.
Peirce, C. S. "Some Consequences of Four Incapacities." In C. Hartshorne and P. Weiss (Eds.), Collected Papers of Charles Sanders Peirce. Vol. 5. Cambridge, Mass.: Harvard University Press, 1936. (Originally published 1868.)
Putnam, H. "The Analytic and the Synthetic." In H. Feigl and G. Maxwell (Eds.), Minnesota Studies in the Philosophy of Science. Vol. 3. Minneapolis: University of Minnesota Press, 1962.
Quine, W. V. O. Word and Object. Cambridge, Mass.: M.l.T. Press, 1960.
Rosenberg, A. "The Supervenience of Biological Concepts." Philosophy of Science, 1978, 45, 368386.
Shimony, A. "Statistical Inference." In R. G. Colodny (Ed.), The Nature and Function of Scientific Theories. Pittsburgh: University of Pittsburgh Press, 1970.
Shweder, R. A. "Likeness and Likelihood in Everyday Thought: Magical Thinking in Judgements About Personality." Current Anthropology, 1977, 18, 637648; reply to discussion, pp. 652 658.
Shweder, R. A. "Rethinking Culture and Personality Theory. Part I."Ethos, 1979a, 7, 255278.
Shweder, R. A. "Rethinking Culture and Personality Theory. Part II."Ethos, 1979b, 7, 279311.
Shweder, R. A. "Rethinking Culture and Personality Theory. Part III. " Ethos, 1980, 8, 6094.
Simon, H. A. "A Behavioral Model of Rational Choice." In P. Nash, Models of Man. New York: Wiley, 1957.
Simon, H. A. The Sciences of the Artificial. Cambridge, Mass.: M.I.T. Press, 1969.
Simon, H. A. "The Structure of IllStructured Problems." Artificial Intelligence, 1973,4, 181201.
Simon, H. A. Models of Thought. New Haven, Conn.: Yale University Press, 1979.
Tversky, A., and Kahneman, D. "Decision Under Uncertainty: Heuristics and Biases." Science, 1974, 185, 11241131.
Valenstein, E. Brain Control. New York: Wiley, 1973.
von Neumann, J. "Probabilistic Logic and the Synthesis of Reliable Organisms from Unreliable Components." In C. E. Shannon and J. McCarthy (Eds.), Automata Studies. Princeton, N.J.: Princeton University Press, 1956.
Wade, M. J. "A Critical Review of the Models of Group Selection." Quarterly Review of Biology, 1978, 53 (3), 101114.
Williams, G. C. Adaptations and Natural Selection: A Critique of Some Current Evolutionary Thought. Princeton, N.J.: Princeton University Press, 1966.
Wimsatt, W. C. "Complexity and Organization." In K. F. Schaffner and R. S. Cohen (Eds.), Proceedings of the Meetings of the Philosophy of Science Association, 1972. Dordrecht, Netherlands: Reidel, 1974.

Wimsatt, W. C. "Reductionism, Levels of Organization, and the Mind-Body Problem." In G. G. Globus, G. Maxwell, and I. Savodnik (Eds.), Consciousness and the Brain: Scientific and Philosophical Strategies. New York: Plenum, 1976a.

Wimsatt, W. C. "Reductive Explanation: A Functional Account." In C. A. Hooker, G. Pearse, A. C. Michalos, and J. W. van Evra (Eds.), Proceedings of the Meetings of the Philosophy of Science Association, 1974. Dordrecht, Netherlands: Reidel, 1976b.

Wimsatt, W. C. "Reduction and Reductionism." In P. D. Asquith and H. Kyburg, Jr. (Eds.), Current Problems in Philosophy of Science. East Lansing, Mich.: Philosophy of Science Association, 1979.

Wimsatt, W. C. "Randomness and PerceivedRandomness in Evolutionary Biology." Synthese, 1980a, 43 (3), 287329.

Wimsatt, W. C. "Reductionistic Research Strategies and Their Biases in the Units of Selection Controversy." In T. Nickles (Ed.), Scientific Discovery. Vol. 2: Historical and Scientific Case Studies. Dordrecht, Netherlands: Reidel, 1980b.