# WILDLIFE SCIENCE: GAINING RELIABLE KNOWLEDGE

H. CHARLES ROMESBURG, Department of Forestry and Outdoor Recreation, Utah State University, Logan, UT 84322

*Abstract:* Two scientific methods called induction and retroduction form the basis for almost all wildlife research. Induction is used to establish reliable associations among sets of facts, whereas retroduction is used to establish research hypotheses about the fact-giving processes driving nature. A 3rd scientific method, the hypothetico-deductive (H-D), is a means for testing research hypotheses, i.e., for gauging their reliability. The H-D method is rarely used in wildlife science. Instead, research hypotheses are proposed, and either made into a law through verbal repetition or lose favor and are forgotten. I develop the thesis that wildlife research should use the H-D method to test research hypotheses, using the threshold-of-security hypothesis for winter mortality for illustration. I show that persistent confusions about the definitions of concepts like carrying capacity, correlation and cause-and-effect, and the reliability of knowledge gained from computer simulation models stem from either inadequate or misused scientific methods.

***J. WILDL. MANAGE. 45(2):293–313***

Like the Kaibab deer herd, progress in wildlife science may be headed for a crash under the weight of unreliable knowledge. Knowledge, the set of ideas that agree or are consistent with the facts of nature, is discovered through the application of scientific methods. There is no single, all-purpose scientific method; instead, there are several, each suited to a different purpose. When the set of scientific methods is incomplete, or when one method is used for a purpose better fit by another, or when a given method is applied without paying strict attention to the control of extraneous influences, then these errors of misuse cause knowledge to become unreliable.

Unreliable knowledge is the set of false ideas that are mistaken for knowledge. If we let unreliable knowlege in, then others, accepting these false laws, will build new knowledge on a false foundation. At some point an overload will occur, then a crash, then a retracing to the set of knowledge that existed in the past before the drift toward unreliability started. Every field that loses quality control over its primary product must undergo this kind of retracing if it is to survive. Of course, some unreliable knowledge inevitably creeps in—a researcher makes a systematic error here, or fails to do enough replications there. All science is prone to human error, and minor retracing continually occurs. But I think part of wildlife science's knowledge bank has become grossly unreliable owing to the misuse of scientific methods, and major retracing is inevitable.

I read published dissatisfactions on seemingly isolated topics as being symptomatic of past misuses of scientific method, e.g., Chitty's (1967) and Eberhardt's (1970) complaints over the continued confusion between correlation and cause-and-effect, Bergerud's (1974) case against the reliance on induction to generalize laws to the exclusion of testing research hypotheses, Hayne's (1978) dissatisfaction with poor experimental designs, Krebs' (1979) frustration with virtually every aspect of small mammal ecology, Caughley's (1980) claims that most large mammal studies "coalesce into an amorphous mass of nothing much" and that white-tailed deer (*Odocoileus virginianus*) and *Drosophila* are the most studied and least understood of animals, and Eberhardt's (1975) skepticism about the predictive value of computer simulation models of ecological systems.

What are these misuses of scientific method? Of the 3 main scientific methods used in virtually all fields, i.e., (1) induction, (2) retroduction, and (3) hypothetico-deductive (H-D), wildlife science uses the 1st and 2nd methods but almost never the 3rd. Induction and retroduction, by themselves, are inadequate for discovering some kinds of knowledge. Instead of realizing this limitation, wildlife science routinely stretches induction and retroduction beyond their limitations as knowledge-finding tools and unreliable knowledge results.

Let me show how this occurs by explaining each method. The method of induction (Hanson 1965, Harvey 1969) is useful for finding laws of association between classes of facts. For example, if we observed over many trials that the amount of edge vegetation in fields was positively correlated with an index of game abundance, we would be using induction if we declared a law of association. The more trials observed, the more reliability we'd attribute to the law. The method of retroduction (Hanson 1965) is useful for finding research hypotheses about processes that are explanations or reasons for facts. For example, if we observed birds caching seeds more on south slopes than on north slopes (facts), and our best guess for the reason of this behavior (our research hypothesis) was that south slopes tended to be freer of snow than north slopes, we would be using the method of retroduction to generalize a research hypothesis about a process providing a reason for the observed facts of bird behavior. The method of retroduction is the method of circumstantial evidence used in courts of law. Retroduction is not always reliable, because alternative research hypotheses can often be generated from the same set of facts.

The H-D method (Popper 1962, Harvey 1969) complements the method of retroduction. Starting with the research hypothesis, usually obtained by retroduction, predictions are made about other classes of facts that should be true if the research hypothesis is actually true. To the extent that experiment confirms or rejects the predicted facts, the hypothesis is confirmed or rejected. Thus, the H-D method is a way of gauging the reliability of research hypotheses acquired by other means.

Wildlife science's workhorse is the method of induction. I believe it is used in a way that gives reliable knowledge. However, induction has a limitation: it can only give knowledge about possible associations between classes of facts. Although this is undoubtedly useful for decision making (e.g., the correlation between a fish's weight and its length is a money-saving association), it cannot give knowledge about the processes that drive nature. Consequently, you can use induction repeatedly without diminishing the question "Why?" When we ask "Why?" we are asking for an explanation, an abstract process that provides a reason for the facts. If the human mind didn't beg for reliable explanations, the method of induction would suffice. That not being the case, the method of retroduction was invented. It is reliable enough to be used in courts of law but, by itself, it is not reliable enough for science. Science has the most stringent standards of all endeavors. If courts of law followed science's strict standards, suspects identified through retroduction would be set free, and their guilt decided in accordance with whether or not the life of crime predicted for them turned up in future facts. That is, the courts would test a retroductively derived hypothesis using the H-D method.

Because wildlife science hardly uses the H-D method, it is stuck with no way of testing the many research hypotheses generated by retroduction. Herein lies the main cause of unreliable knowledge. The research hypotheses either are forgotten, or they gain credence and the status of laws through rhetoric, taste, authority, and verbal repetition. Leopold's (1933) book *Game Management* lists 9 entries under "hypothesis"; I think none has ever been tested by the H-D method. Errington's (1945) threshold-of-security hypothesis, a hypothetical process of winter mortality, is often stated as a law, but it is a retroductively derived hypothesis, and it, strictly speaking, remains untested.

The normal pattern of university graduate and faculty research—spending hundreds of hours watching, describing, and quantitatively recording the habits of animals, relating their habits to environmental facts, analyzing the data using a computer and contemporary statistical analysis, and then drawing conclusions from patterns in the summarized data—produces reliable knowledge to the extent that induction and retroduction, properly used, will allow. But for me the reliable parts, inductively derived correlations about events, are often not interesting or even useful, whereas the interesting parts, the retroductively derived reasons for what is going on, are often unreliable speculation. The H-D method is a way of raising the reliability of this speculation and, hence, the overall reliability of our knowledge. It is not a cure-all. It cannot suggest good questions for research. It cannot be used to test every conceivable research hypothesis, for reasons of experimental costs and lack of creativity on the part of researchers. It can be misused like any other method of science, but can also lead to the discovery of reliable knowledge about processes.

The remainder of this paper will (1) explain the H-D method in detail; (2) show how the H-D method could be used to test Errington's (1945) threshold-of-security hypothesis; (3) show why the kind of general-purpose data routinely collected by game agencies is inadequate for testing research hypotheses; (4) show how an understanding of the H-D method resolves persistent confusions in wildlife science thought; and (5) contrast science with planning.

I thank D. Anderson, R. Brown, W. Clark, F. Wagner, and M. Wolfe, Utah State University, and the referees for their comments.

## ESSENTIALS OF THE H-D METHOD

Terms critical to understanding the H-D method must first be defined: viz., theory, research hypothesis, statistical hypothesis, and test consequence. The term *theory* means a broad, general conjecture about a process. For example, the Lotka-Volterra competition equations (Emlen 1973) represent a theory about the process of competition between 2 animal species. A *research hypothesis* is a theory that is intended for experimental test; it has the logical content of the theory, but is more specific because, for example, the location and animal species must be specified. A research hypothesis must be tested indirectly because it embodies a process, and experiments can only give facts entailed by a process. The process itself is abstract, removed from the senses, and nonfactual. The indirect test is conducted by logically deducing 1 or more *test consequence(s),* i.e., predicted facts, such that if the research hypothesis is true, then the test consequence(s) must be true, and the test consequence(s) must correspond to a feasible experiment, e.g., one that is not technologically impossible or so costly as to be impractible.

For example, consider the question of

how salmon find their way upstream to their home spawning grounds. The answer, "salmon navigate by vision alone," is a research hypothesis ($H$), i.e., a conjecture about a process of navigation. A test consequence ($C$) is "a group of salmon that has been captured and blinded as they begin their upstream migration will not reach their home tributary spawning grounds in numbers greater than expected by chance, whereas a nonblinded control group of equal size that was spawned in the same tributory as the blinded fish will return to their tributary in numbers greater than expected by chance." The fact of the test consequence $C$ must then be obtained by experiment, e.g., tagging smolts before their migration to the lake or ocean, recapture of those returning to spawn, and subsequent recapture of blinded and control-group salmon after they have swum upstream.

The determination of whether or not $C$ is true or false by reference to experiment requires a *statistical hypothesis* to be tested, e.g., the null hypothesis $H_0$: "control and blinded salmon return in equal numbers." Thus, a research hypothesis is a conjecture about a process, whereas a statistical hypothesis is a conjecture about classes of facts entailed by the process. In general, alternative test consequences can be used to test a research hypothesis. For example, an alternative test consequence is "when ink is metered into the stream so that vision is totally impaired, the fish will not reach their spawning tributary in numbers greater than expected by chance."

Because a test consequence prescribes the experiment necessary to ascertain the truth or falsity of $C$, the H-D method demands creative thinking. Creative researchers will search for test conclusions that require experiments beset by minimal statistical noise, that are cheap to perform, and that allow tight control of extraneous influences (note that the 2nd test conclusion is not as good as the 1st, because it doesn't allow for a control group). Successful researchers are defined, in part, as those who make a career of choosing the right trade-offs between these usually conflicting considerations.

The experiment's outcome determines whether $C$ is judged to be true or false. If $C$ is true, then $H$ can be either true or false, and we say that the evidence supports or confirms the truth of $H$, i.e., is consistent with $H$ being true. For example, consider the hypothetical limiting case in which somehow the truth or falsity of $C$ is known with certainty, i.e., no test of a statistical hypothesis is required. If $C$ turns out to be true, i.e., fewer blinded than nonblinded return, then support for the conjecture $H$ that "salmon navigate by vision alone" is evidenced. Further, the more replications carried out with the same outcome, the stronger the support is, although the truth of $H$ can never be declared with certainty because it is possible, for example, that $H$ might really be false but other factors, such as a propensity for blinded fish to die, could be making $C$ true.

On the other hand, if $C$ turns out to be false, then $H$ is false, provided that none of the background conditions required to make $H$ entail $C$ are violated. For example, if $C$ is false, i.e., blinded and nonblinded return home in equal numbers, then $H$ is false provided that $H$ really does entail $C$. If blinded fish exhibit a schooling behavior not dependent on vision and get home by tagging along behind sighted fish, then, of course, $C$ being false is not justification for the statement that $H$ is false. An experimenter can never gain complete assurance that the statement "the truth of $H$ entails the truth of $C$" is true. Thus, even $C$ being false does not provide complete assurance that $H$ is false. However, the more certain a researcher is that the background conditions are indeed true, the more certain he will be in pronouncing $H$ to be confirmed when $C$ is true, and $H$ to be falsified when $C$ is false.

The details of the H-D method that fill out this brief outline are covered by Popper (1962), Platt (1964), Baker and Allen (1968), Harvey (1969), Medawar (1969), and Rachelson (1977). Bergerud (1974) used the H-D method to design a hypothesis test about the processes that cause caribou (*Rangifer tarandus*) populations to decline.

Unless otherwise noted, the words "research hypothesis" and "test consequence" have been shortened to "hypothesis" and "consequence," and the symbols $H$ and $C$ used to signify them.

## TESTING THE THRESHOLD-OF-SECURITY HYPOTHESIS

In this section I show how the H-D method can be used to test Errington's (1945) threshold-of-security hypothesis. I have 2 reasons for including this, even though it is not essential to the central arguments. First, this hypothesis is well known to virtually everyone in the wildlife science and management professions, and is a star example of the incorrect use of the method of retroduction to make a law out of what is really a research hypothesis. Three decades have passed sine Errington formulated it, and since that time it has passed into the profession's lore without being tested or carefully thought out. Second, merely stating that it can be tested using the H-D method is not enough. The details need to be spelled out to show that testing is feasible.

### Threshold Hypotheses and Consequences

Errington (1945), in a study of processes regulating bobwhite (*Colinus virginianus*) survival over winter in the vicinity of Prairie du Sac, Wisconsin, 1929–44, formed the view, using induction and retroduction, that winter mortality could be governed by a threshold process that could be considered approximately constant from year to year. This constant-threshold hypothesis is stated in many places in the literature; Wagner (1969:268–269) puts it this way: "[the hypothesis] . . . visualizes game populations occupying environments with limited and generally well-fixed capacity to protect animals during the winter season. Each year the reproductive season produces a number in excess of the winter threshold. This annual surplus inevitably disappears through predation, weather, or emigration because of the animals' intolerance to crowding into the limited habitat niches. The animals living within the security threshold experience little if any losses, barring catastrophic weather incidents." He goes on to state the management consequences of the hypothesis: "If pitched so as to remove no more than a number equivalent to the annual surplus, hunting can take a portion of animals without increasing the fall-spring mortality rate or affecting the population level." Thus, when the take does not exceed the annual surplus, harvest mortality is 100% compensatory with natural mortality.

A constant thresold was the initial impression formed by Errington. He later changed to that of a variable threshold and described it this way (Errington

1945:11–12): "Except in the event of emergencies, populations living below threshold values wintered with slight reduction through predation and self adjustment. If exceeding thresholds, populations betrayed instability and pronounced vulnerability to predation until again reduced to secure levels." He commented on variation in the threshold: ". . . I think that for many years I was seriously misled in my attempts to analyze threshold phenomena by the evidences of year-to-year constancy in threshold values . . . ," and concluded that data taken after the winter of 1935–36 supported a variable threshold concept. Errington didn't identify the management consequences of a varying threshold. It is clear, however, that a threshold that varied in an uncertain manner would create an annual surplus that could not be identified with certainty before winter mortality occurred and therefore could not be easily exploited.

These 2 hypotheses are as follows. Letting $H_c$ and $H_v$ stand for the constant threshold and the variable threshold hypotheses, respectively:

$H_c$ = For a given area and species, the number of animals surviving fall to spring can be no greater than a threshold value. This threshold accounts for all forms of natural mortality, barring catastrophic weather events, and is constant from year to year.

$H_v$ = For a given area and species, the number of animals surviving fall to spring can be no greater than a threshold value. This threshold accounts for all forms of natural mortality *including* catastrophic weather events. The threshold value realized in any given year is probabilistic, and therefore cannot be predicted accurately before its occurrence.

$H_c$ bars catastrophic weather events as a component of natural mortality, because their inclusion would, by definition, destroy the constancy of the threshold. Similarly, in stating the varying-threshold view, Errington (1945:11–12) bars "emergencies," i.e., catastrophic events. However, no generality is lost in the analysis if catastrophic weather events are incorporated as a component of overall variability. The statement of $H_v$ allows this more general and more realistic view.

## Mathematical Statement of $H_c$ and $H_v$

We assume that a population exists in an area which experiences a period of uncertain natural mortality from fall to spring (other bottlenecks like drought-induced summer mortality could be treated analogously).

Let $x$ = fall population size, $y$ = spring population size, $w$ = number dying during winter, and $t$ = threshold of security.

The conservation of population over the period requires that

$$y = x - w. \tag{1}$$

The threshold $t$ is defined as the maximum population size that can survive the period. If $x$ is greater than $t$, the excess $x - t$, i.e., annual surplus, is lost through natural mortality. Thus, the number dying in the period is

$$w = \begin{cases} 0 & \text{if } x < t \\ x - t & \text{if } x \geqslant t. \end{cases} \tag{2}$$

Using Eq. (2) in (1) gives the spring population size $y$ as a function of the fall population size and threshold:

$$y = \begin{cases} x & \text{if } x < t \\ t & \text{if } x \geqslant t. \end{cases} \tag{3}$$

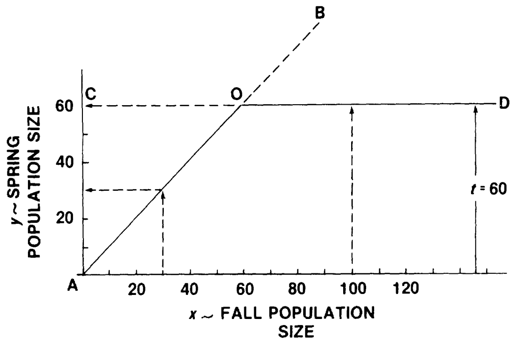


Fig. 1. Spring population size $y$ related to fall population size $x$ when the winter threshold of security $t$ = 60.

When $t$ is considered to be constant and known, (3) is the mathematical statement of $H_c$. When $t$ is probabilistic, i.e., a random variable, (3) is the mathematical statement of $H_v$.

## Management Consequences of $H_c$

Equation (3) (with $t$ = 60) is graphed in Fig. 1: line AB is $y = x$; line CD is $y = t$; the function relating $y$ to $x$ is line AOD. Lauckhart and McKean (1956:61) give an apt "bucket" analogy for the logical consequences of $H_c$, which Fig. 1 mirrors. Regarding stocking of pheasants (*Phasianus cochicus*) in the fall after harvest, they say, "To plant birds in an already 'full' habitat is like pouring water into a full bucket." They go on, "If hens are going to die down to a certain capacity number, it should be possible to allow the hunter to take some . . . ." In Fig. 1, line CD is the bucket's lip. It exerts its effect when $x \geqslant t$. For example, when $x$ = 100, reading from line CD gives a spring population size $y$ = 60. The annual surplus $x - t = 100 - 60 = 40$ is a so-called "doomed surplus." Consequently, it can be harvested before winter without affecting $y$, i.e., $y$ is destined to be 60 with or without harvesting these 40 animals.

On the other hand, when $x < t$, the bucket is not yet full. In this case, animals harvested will reduce, in 1-to-1 fashion, the spring population size $y$. If no harvest is made, all fall animals are carried over to spring (shown in Fig. 1: when $x$ = 30, reading from line AOB, it is clear that $y$ = 30). Finally, if $x$ = 30 and more animals are added to the population in the fall, then it is clear that an addition of $t - x = 60 - 30 = 30$ is the maximum permitted without spilling over the bucket's lip. If more than 30 are added, the excess over 30 will become the annual surplus.

## Management Consequences of $H_v$

Consider now the varying-threshold case, $H_v$. The situation can be pictured analogously as a stack of graphs, perhaps several thousand in number, similar to Fig. 1, except that the threshold value $t$ is different on each graph. Each graph represents 1 winter season on the same area; the stack represents several thousand winter seasons. Suppose these graphs were used in a simulation exercise with a game manager in the following way. The manager would be given complete information about the frequency of occurrence of different values of $t$ in the stack, and could therefore calculate the mean threshold E($t$) (expected value of the random variable $t$) and the standard deviation of the threshold SD($t$). The manager would pretend that a specified fall population size, say $x$ = 100 for the discussion, existed on the area every fall. The manager's objective would be to pick a harvest level $h$ that would take exactly the annual surplus. After the manager decided upon a harvest level, a graph would be drawn at random from the stack to represent nature's choosing of the winter threshold.

It is possible that on given trials the

manager would achieve the objective exactly. For example, suppose $h = 20$ is decided, and $t = 80$ occurs. This "steals" the annual surplus ($100 - 80 = 20$); no more, no less. Or, if $h = 0$ is decided and $t = 160$, the annual surplus, which is 0, is, correctly, not harvested. Chance, however, makes it so that over a period of trials the manager could not on average exactly remove the annual surplus. For example, some trials will be of the form $h = 20$, $t = 200$, i.e., the annual surplus was 0, but 20 animals were harvested; the consequence of this error is to reduce the spring population $y$ by 20 animals. A 2nd error is of the form $h = 20$, $t = 40$, i.e., the annual surplus was $100 - 40 = 60$, but $60 - 20 = 40$ of these animals were not harvested and therefore remained as annual surplus. Thus, not taking the full annual surplus or, conversely, taking more than the annual surplus are possible consequences under $H_v$. The problem is as difficult as deciding how much milk can safely be poured into a bucket without spilling when the decision must be made before the bucket is seen, and the only information available is that the height of the bucket's lip varies according to a probability distribution with known mean and standard deviation.

Mathematics can aid the understanding of this complex situation. A useful mathematical index is the mean value of the spring population size $y$ for a given value of the fall population size $x$. This is denoted by $E(y|x)$, which reads "the expected value of $y$ given $x$." Its meaning is as follows: if the fall population size was maintained at the level $x$ (where $x$ is a specified number of animals) every year, and if the value $t$ of the threshold occurred each year in accordance with a specified frequency, i.e., a probability density function, then the average over the years of all the possible resulting $y$ values would be $E(y|x)$. A 2nd statistic of interest is the standard deviation of $y$ given any value of $x$, denoted by the symbol $SD(y|x)$. Thus, $E(y|x)$ is a measure of the central tendency of the spring population size, whereas $SD(y|x)$ is a measure of its precision.

In $H_v$, the threshold $t$ is a random variable. To derive $E(y|x)$ and $SD(y|x)$, the probability density function (p.d.f.) of $t$, i.e., $f(t)$, must be assumed. The choice of $f(t)$ made here is the gamma distribution

$$f(t) = \frac{1}{b^a\Gamma(a)} t^{a-1} e^{-t/b}, \quad \text{where} \quad 0 \leq t < \infty. \tag{4}$$

Whether or not this form for $f(t)$ portrays reality is not an important issue; the choice is made so that numerical results can be obtained simply, and the general conclusions drawn would be the same for any distribution.

Two sets of parameters $a$ and $b$ are used, and the density functions, denoted as Cases 1 and 2, take the forms shown in Fig. 2. Both have identical means of $E(t) = 60$, but Case 1 has a larger standard deviation. The functions are bell-shaped but skewed, and it follows that extremely large or small values of $t$ are less likely than values in the middle range, say $30 \leq t \leq 90$.

With $t$ a random variable and $y$ related to $t$ by eq. (3), $y$ itself is a random variable with p.d.f. $g(y)$. The required transformation from $f(t)$ to $g(y)$ is easily obtained. Let the function (3) relating $y$ to $t$ be denoted as $y = h(t)$. The transformation in the range $0 < t < x$ is given by

$$g(y) = f[h^{-1}(y)] \frac{dh^{-1}(y)}{dy} \tag{5}$$

(for example, see Freund 1962:132–137).

In the range $x < t < \infty$, eq. (3) gives $y = x$. So values of $t$ map into a probability mass $m$ at $y = x$. The value of $m$ plus

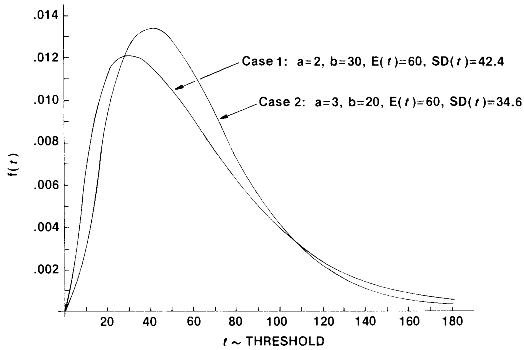


Fig. 2. The gamma distribution $f(t)$ representing the probability density function of the winter threshold of security $t$. The 2 cases differ by parameter settings $a$ and $b$; both have the same mean $E(t) = 60$, but different standard deviations $SD(t)$.

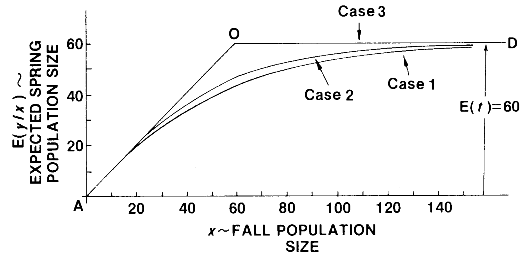


Fig. 3. Expected spring population size $E(y|x)$ as a function of the fall population size $x$. Cases 1 and 2 are based on the gamma-distribution p.d.f.'s shown in Fig. 2. Case 3 is based on a degenerate p.d.f. with mean $E(t) = 60$, but $SD(t) = 0.0$.

the integral of $g(y)$ from 0 to $x$ must, by definition, be 1.0, and this gives a means for computing $m$:

$$m = 1.0 - \int_0^x g(y)\,dy, \tag{6}$$

with $g(y)$ given by eq. (5).

The mean and variance of $g(y)$ are conditional upon $x$:

$$E(y\,|\,x) = \int_0^x yg(y)\,dy + xm \tag{7}$$

$$\mathrm{Var}(y\,|\,x) = \int_0^x y^2 g(y)\,dy + x^2 m - E(y\,|\,x)^2 \tag{8}$$

where the 1st 2 terms of (8) are $E(y^2|x)$. The standard deviation $SD(y|x)$ is the positive square root of $\mathrm{Var}(y|x)$.

The p.d.f. $f(t)$, given by (4) for Cases 1 and 2 (Fig. 2), is transformed to $g(y)$ by using eq. (5). The value $m$ is computed from $g(y)$ by using eq. (6). $E(y|x)$ and $\mathrm{Var}(y|x)$ are calculated using $g(y)$ and $m$ in eqs. (7) and (8), and the results are graphed in Figs. 3 and 4.

These graphs are based upon 3 forms of the p.d.f. $f(t)$, all having the same mean $E(t) = 60$. They differ, however, in standard deviation $SD(t)$: Case 1 has the largest, $SD(t) = 42.4$; Case 2 has the next largest, $SD(t) = 34.6$; Case 3 has the theoretically smallest, $SD(t) = 0.0$. The p.d.f.'s for Cases 1 and 2 are shown in Fig. 2; Case 3 cannot be shown because it is the limiting case when all of the probability is concentrated at the value $t = 60$.

In the limit as $SD(t)$ tends to 0.0, $H_v$ becomes $H_c$. This is clearly shown in the comparison of Figs. 1 and 3. Figure 1 is based upon $t = 60$, and line AOD is the function relating $y$ to $x$. Similarly, Fig. 3, Case 3, is based upon $E(t) = 60$, and line AOD is the function relating $y$ to $x$. Because lines AOD are the same in both

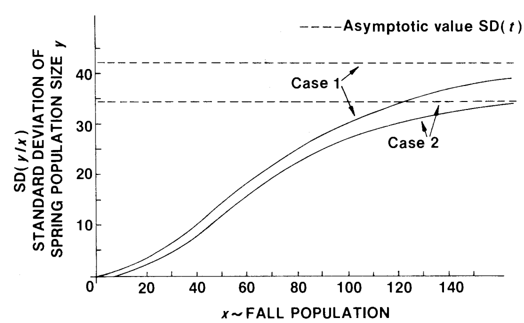


Fig. 4. Standard deviation of spring population size $SD(y|x)$ as a function of the fall population size $x$. Cases 1 and 2 are based on the gamma-distribution p.d.f.'s shown in Fig. 2.

figures, $H_v$ is identical to $H_c$ when $SD(t) = 0.0$. Thus, the reader should think of Case 3 in Fig. 3 as the consequence of $H_c$, whereas the other 2 cases are the consequences of $H_v$.

Figure 3 shows that a varying threshold (either Cases 1 or 2) makes, on average, 100% compensation between natural mortality and harvest an impossibility. The only place on Fig. 3 where reducing the fall population size $x$ through harvest will not change the expected spring population size $E(y|x)$ is the constant-threshold Case 3 in the particular event that the value of $x$ after harvest is 60 or larger. Moreover, when $x$ is 60 or less before harvest, 0% compensation between natural mortality and harvest mortality must occur under Case 3.

The effect of a varying threshold is to prohibit the extremes of 0 and 100% compensation between mortality and harvest. This is clearly shown in Fig. 3. For 0% compensation to occur, the function relating $E(y|x)$ to $x$ must make a 45° angle with the $x$-axis at the origin; for 100% compensation to occur, the function must lie parallel to the $x$-axis. Case 3 is the only case that meets these requirements, although in the limit as $x$ tends to 0 the other cases approach 0% compensation, whereas in the limit as $x$ tends to $\infty$ the other cases approach 100% compensation.

A real situation in which a constant threshold was operative would be simple to manage. In such a case, the decision to remove the annual surplus each year nearly dictates itself. A varying threshold scenario, however, introduces a trade-off between a given level of harvest and an expected decrease in the spring population base. This decision does not dictate itself, but, to the contrary, depends upon the valuations placed upon the components to be traded off, valuations that differ across species and area.

The standard deviation of $y$, $SD(y|x)$, is graphed in Fig. 4 for cases 1 and 2. It serves as an index of uncertainty in predicting $y$ for the following spring. As Fig. 4 shows, $SD(y|x)$ increases in sigmoidal fashion with $x$. This makes predictions more uncertain when $x$ is large than when small. There is an upper limit to the level of uncertainty; $SD(y|x)$ can never exceed $SD(t)$, the standard deviation of the threshold. Thus, $SD(y|x)$ depends upon an uncontrollable component ($SD[t]$) as well as on a controllable component ($x$). A manager can gain a more precise prediction of the spring population by reducing $x$ through harvesting, but this will carry with it the consequence that $E(y|x)$ will also be reduced.

This analysis shows that a constant threshold yields greater management benefits than a varying threshold, because 100% compensation of harvest and natural mortality is possible and because harvesting the annual surplus is without the consequence of altering future population size.

## H-D Method Used to Test Threshold-mortality Hypotheses

Because $H_c$ is a special limiting case of $H_v$, the discussion will center on testing $H_v$ for a given population in a defined area. We would select smaller study sites within the larger area, each stocked in the fall at a different level $x$, covering at least the range bounded by the historic low and high levels. The spring population level $y$ would be subsequently estimated on each area. If a graph of $y$ vs. $x$ followed the shape of curve AOD in Fig. 1, we would have confirming evidence that a winter threshold of security was operating. If a repeat of this experiment over a period of years produced a curve shaped like Fig. 1 each year, but the threshold $t$ varied, then this would be confirming evidence for $H_v$. If instead $t$ was invariant, this would confirm $H_c$. Finally, if in any year the shape of the experimentally drawn curve departed from the characteristic shape of Fig. 1, then both $H_c$ and $H_v$ would be false.

The procedure uses the curve AOD in Fig. 1 as a test consequence. It is crucial that requisite background conditions not be violated, i.e., sources of experimental error must be invariant to the extent that if either hypothesis is in fact correct, then experimentally drawn curves such as AOD will emerge to within a defined tolerance. Important sources of confounding error occur when any of the following is true: (1) study areas in a given year are not the same with respect to the forces causing winter mortality, (2) the experimental animals are not the same across study areas and years with respect to their resistance to the forces of winter mortality, (3) inadequately estimated emigration or immigration after stocking occurs, and (4) $y$ or $x$ cannot be estimated on each study area with high precision and small bias. To the extent these sources of error are not eliminated, experimentally drawn functions will likely depart from Fig. 1 when the hypotheses tested are actually true. Conversely, the hypotheses could actually be false but errors might "bend" the true relation of $y$ and $x$ into a curve like Fig. 1. The experimenter would have only partial control of these sources of error. In the end it would come down to assuming that the remaining sources were self-controlling. Thus, getting conclusive results depends upon the experimenter's care in exercising control, as well as being lucky (unfortunately the state of luck will never be known).

In this section I have speculated on the prognosis for designing conclusive experimental tests. I conclude that uncontrolled errors make for marginally conclusive tests at the present time. Two influences can change this for the better: the discovery of test consequences that are less beset by error or that offer more control of sources of error, and changes in technology, such as more precise population estimation methods.

## Uses of General-purpose Data

I call the kind of data routinely collected by game agencies "general-purpose data." I ask whether they have scientific uses. Can they be used with the method of induction, method of retroduction, and the H-D method to discover reliable knowledge? The answer is yes and no.

I can see the method of induction being used with these data to obtain reliable laws, e.g., the correlation of the number of dead fawns with snow depth and duration. On the whole, however, I see unreliable knowledge resulting when the method of retroduction or the H-D method is used with these data. To understand a process of interest, the process must be isolated from other processes by exacting experimental control. However, general-purpose data are not collected under controlled conditions.

Used with general-purpose data, the method of retroduction contains the flaw of incorporating the effects of unknown factors into the derived research hypotheses. Similarly, the H-D method can only produce reliable knowledge when background conditions are held to a tight tolerance. If the tolerance is lost, then the researcher will probably conclude something that in essence is more a result of error than substance. It goes beyond reasonable doubt for researchers to assume that nature delivers tightly controlled experiments without prompting. The creation of knowledge of processes from general-purpose data is therefore suspect.

History illustrates the pitfalls of loosely applying the H-D method to general-purpose data. Lauckhart (1955) and Lauckhart and McKean (1956) interpreted data from pheasant population studies as supporting the threshold-of-security hypothesis, but the pheasant population data of Wagner et al. (1965) and Wagner and Stokes (1968) were interpreted as not supporting the hypothesis. Who can say that unknown factors are not giving conclusive results when there either are none or truly conclusive results are being obscured by errors?

## OTHER PROBLEMS WITH WILDLIFE SCIENCE

Wildlife science has other problems befitting analysis. I will cover problems with concept definition, confusion between cause-and-effect and correlation, use of slipshod experimental controls, and the fixation on statistical methods.

### Problems with Concept Definition

Key wildlife science concepts suffer from multiple and unclear definitions. For example, nearly 3 decades ago Edwards and Fowle (1955) concluded that more than a dozen different meanings of the concept "carrying capacity" were in use, and that most were vague and almost meaningless. They tried to right the situation by proposing a new, clearer definition. They failed; the confusion is undiminished today.

There is a mistaken belief that a profession sets the meanings of its concepts by decree. To be accepted, a concept must have appeal, and it can gain the necessary appeal in 2 ways. First, the concept can function in an inductively established law. For example, if it could be shown that a given definition of carrying capacity entered into inductively established laws with other concepts such as time, then the concept would gain appeal. Second, the concept can function in a law established by the H-D method. For example, if a given definition of carrying capacity functioned in a theory with other concepts, and if the theory became law through experimental test, then the concept would gain appeal.

The history of science shows that most of the concepts with staying power are those that function in laws established by the H-D method. For example, the concept of mass is substantiated by Newton's and Einstein's laws. When wildlife science decides to propose and test theories built around different concepts of carrying capacity, then the correct concept will emerge in those theories that pass experimental muster. When a science has no way of telling when a theory and the concepts it integrates are in error, then it has no way of telling which concepts are right.

### Cause-and-effect vs. Correlation

One of the aims of wildlife science is to find cause-and-effect relationships among variables, for when cause-and-effect is found, then control may be possible. To say that a change in variable A causes a change (effect) in variable B, i.e., that B depends on A, requires a particular experiment: we merely introduce a change in A and see whether a change in B follows. If this occurs, and if in a control unit or group the variable A was not changed and a corresponding change in B did not follow, then we have evidence that A causes B. If A unerringly causes B over many trials, then at some point the method of induction leads to the pronouncement of a causal law.

This is the basic method used in clinical experiments in medical research (Feinstein 1977:17–70). The variables A and B can be either binary or continuous. An example using binary variables is: A is a treatment variable taking on the states of (1) treatment applied or (2) treatment not applied, and B is an effect variable taking on states of (1) an effect is observed and (2) no effect is observed. An example using continuous variables is: A is the energy (kcal/kg) in a diet fed to deer and B is the change in body weight (kg); Verme and Ozoga (1980) performed this type of study for white-tailed deer fawns under well-controlled conditions.

Chitty (1967) has proposed a similar but weaker method for binary variables. It differs from the method above in that nature selects the 2 settings of the supposed cause. Nature often won't yield control in field studies, so this design has appeal. The weakness, however, is that the data can follow the pattern required for saying A causes B, yet if we could seize control of A ourselves we might find that varying A produced no effect on B, i.e., the supposed cause-and-effect relation could stem from a 3rd variable that caused A and B to covary.

The scientific literature contains many examples of correlations between 2 continuous variables A and B being interpreted as one variable causing the other. For example in wildlife science, Eberhardt (1970) took 3 articles to task for inferring the existence of density-dependent regulation of populations from correlations. Bergerud (1974) questioned the logic in 2 articles in which observed correlations between fecal-pellet counts and the abundance of lichens were used to suggest that caribou require lichens for survival. And, Wagner (1978:198) used a correlation between the annual instantaneous rate of change in coyote (*Canis latrans*) populations and a black-tailed jack rabbit (*Lepus californicus*) population index to conclude that jack rabbit numbers are a determinant of long-term coyote density, and concluded that "if the mean jack rabbit densities over a period of years were increased, annual rates of change in coyote numbers would be largely positive for the years immediately following."

Let us examine this last example in some detail. Clearly, there is a better but more expensive way to test the hypothesis that jack rabbit numbers are a determinant of long-term coyote density. The hypothesis would be supported if we took control of rabbit numbers and showed that changes made in levels of stocking were accompanied (with suitable time lags) by changes in coyote density. Failure to observe the effect, i.e., if coyote density followed no pattern as we changed stocking, would (if it could not be explained away as statistical noise) disprove the hypothesis.

On the one hand, we have an expensive experiment that can give reliable knowledge and a less expensive experiment involving correlations. Which is the better approach? Reliable and expensive? Or less reliable and less expensive? Depending on the costs, the case can be made either way. It is obvious, for example, that medicine could, at high moral and social costs, design experiments capable of producing reliable knowledge on the possible cause-and-effect link between cigarette smoking and lung cancer. But the costs are too high, and therefore medicine is restricted to making less reliable statements based on correlations observed in an uncontrolled setting. It is important to note that they choose this over the alternative of not studying the problem.

Correlation between 2 variables A and B, although a necessary condition for causality, is not sufficient. There is, however, justification for looking for correlations. Correlations offer weak support for making statements using the terms "causes," "determines," or "depends upon," but the history of science is replete with strong pronouncements of cause-and-effect based solely on correlations, only to find that later studies showed no such link.

Consequently, cause-and-effect should not be strongly stated when correlation is the sole evidence. If, however, a correlation is accompanied by other evidence (e.g., other corroborative evidence, the elimination by control of other variables conceivably responsible for the correlation, the demonstration of the correlation under a wide variety of circumstances that allow other possible influences to vary), and logically the dependence makes sense (e.g., the jack rabbit is a staple in the coyote's diet), then support for cause-and-effect is strengthened. Depending on the supporting evidence, the phraseology should range, I think, from "weakly supports," when a correlation is the sole evidence, to something short of "strongly supports," when correlation is accompanied by a variety of independent corroborating evidence. "Strongly supports" should be reserved for direct demonstration of cause-and-effect.

## Thoughtful Use of Experimental Controls

I have selected 2 cause-and-effect studies, one from medicine and the other from wildlife science, to demonstrate the effects of experimental control on gaining reliable knowledge. The medical study (Nelson et al. 1980) tests the research hypothesis that the Leboyer Method of delivering babies in a dark, quiet environment results in the babies growing up to be healthier and calmer. One group of mothers received the Leboyer Method of delivery, the other group a conventional delivery, and the well-being of the babies was assessed for some months afterward and compared. The Leboyer Method had no effect. Thoughtful use of experimental controls greatly increased the reliability of the resultant knowledge. All conceivable alternative determinants of the babies' future health and calmness were controlled: the same obstetrical practice and the same delivery room were used, the group of mothers was selected on the basis of sharing certain common characteristics possibly related to the supposed effect, and the mothers were randomly assigned to the 2 methods of delivery.

Now consider a comparable study on the effects of vegetation interspersion on pheasant abundance (Taylor et al. 1978). Interspersion on a land unit in crops changed over a 20-year period, and an index of interspersion correlated with a pheasant density index. The authors conclude that the relationship is "useful for predicting changes in pheasant density, given anticipated land-use changes." How much more reliable the statement would be if controls had been used, if matched land units similar in all conceivable determinants of pheasant density except edge had been used and the experiment conducted over the same time period. Looked at another way, suppose the medical study had been done the way the wildlife study was, not paying strict attention to controls. It would go something like this: a group of mothers who happened to be handy would be given the Leboyer Method of delivery in different hospitals by different obstetricians; 20 years later other mothers would get conventional deliveries; the well-being of the babies produced by the 2 groups would be compared. If medicine wanted unreliable knowledge, this is how they could go about getting it.

Research needs to be done correctly the first time. For example, Boag and Lewin (1980) tested the efficacy of different objects in deterring waterfowl from using natural and polluted ponds. They conclude their article with an apology for not doing the study under strictly controlled conditions: "... the causal elements in the correlation reported herein need to be tested by using experimental and control ponds in a given time period and thus avoid the necessity of using the same pond successively as a control and then as an experimental pond." I prefer conclusive over nonconclusive studies. Studies saying "here's how we did it; now here's how you can do it right" leave me cold not because they are wrong but because progress is unnecessarily retarded.

The reliability of knowledge is only partially determined by the dedication of researchers. No amount of dedication can make up for lack of experimental controls. In some studies the expense required to achieve control is not worth the expected gain in reliability. But for myself, I would rather see a handful of studies containing highly reliable knowledge than scores of studies containing something less. I would trade all of the studies that have been done on edge and interspersion for one carefully controlled study.

## Fixation on Statistical Methods

In the wildlife literature of the past decade one finds an increasing use of nonparametric and multivariate statistical methods and computer analyses. There can be no doubt that progress in planning and scientific understanding is aided. I have noticed, however, scientific studies that lacked thought and were dressed in quantitative trappings as compensation. It's easy to collect data, perform statistical tests of hypotheses of the "no pattern" variety, and restructure the data using a computer. But all studies must be able to stand up to the question "So what?" I think that too many can't.

The mind must direct research. The processes of upstream salmon navigation are well understood (Hasler 1966, Hasler et al. 1978) because this research was directed at answering specific questions. Every phase of this research—from the initial generation of alternative research hypotheses explaining possible navigational mechanisms to the subsequent tests of these hypotheses using the H-D method—has been guided by a repeating cycle of questions, tentative answers, and tests of research hypotheses. Statistical analyses and the computer played an essential, but secondary, role. The questions and research hypotheses always directed the subsequent quantitative analyses.

Turning the process around and putting the quantitative analysis first is a quantitative natural history study. This can play a vital role at the start of research into an area where little is known by suggesting questions or research hypotheses. This is not bad, but it is time to shift the emphasis to hypothesis testing rather than hypothesis creation; otherwise we'll become swamped with untested ideas. In short, quantitative data analyses that are window-dressing should not be tolerated, those that are natural history studies should be tolerated, and those that play a role in the testing of research hypotheses should be encouraged.

## SCIENCE AND PLANNING

Science and planning are the respective domains of wildlife science and wildlife management. These domains are philosophically distinct, yet because each shares many of the same activities and tools, viz., data collection, statistical methods, and computer simulation models, their differences often pass unnoticed. Yet criticism of the use of common tools is baseless unless these differences in how the tools are used are understood. Accordingly, I will contrast science and planning, and discuss the use of computer simulation modeling in the 2 domains.

### The Nature of Planning

Science and planning are different kinds of decision-making. Science (the H-D method) exposes alternative theories to facts, selects the best theory, i.e., that which agrees closest with fact, and gives it the name "law." Planning exposes alternative images of a future possible world to the decision-maker's values, or preferences, and selects the best image, i.e., that with the highest value. The essential difference is that science uses fact as its standard for selection, whereas planning uses values.

The images in planning are composed of scientific knowledge, common sense, rule-of-thumb knowledge, and theories that are as yet untested, i.e., hunches (Boulding 1956, 1980). Because the image of the status quo will change over time due to influences outside the planner's control, planning is often necessary to counter these uncontrollable influences. Man's imperative to plan is so strong that planning routinely goes on even when scientific knowledge is totally absent from the planner's images. When the imperative to plan takes hold, a planner will enter into the planning process with the best knowledge, tools, and thought at hand, regardless of how imperfect they are. For example, this nation's macroeconomic policy is largely geared to projected images made by computer simulation models of the economy. Yet the elaborate mathematical equations comprising these models represent untested economic theory, and by even the loosest standards of science their predictions fail to agree with economic fact as revealed in the future. Or consider that alternative plans for deployment of land-based intercontinental ballistic missiles are characterized by little scientific knowledge (Feld and Tsipis 1979): the probabilities of destroying the other side's missiles are crude hunches, and the probabilities of how the other side will target its missiles are based on a common-sense image of rational behavior. Finally, a recent wildlife management text (Giles 1978) draws only minimally upon the thousands of scientific articles that have appeared over the years in this journal. Yet no one would argue that planning for the economy, defense, or wildlife should not be undertaken until every part of the images used in planning is substantiated by scientific study.

Science uses relatively absolute and tight tolerances for deciding which theories and hypotheses should be called law. Planning does not use tolerances for deciding what is the best plan, but instead defines "best" as relative to the set of alternative images. Thus, science may never arrive at laws in certain areas for no theory may be within the tolerance for truth, but planning will always arrive at a best plan. Although science and planning share common tools, science and planning have different norms for certifying ideas, and hence criticism of these tools must take into account the domain of their use.

## Computer Simulation Modeling as a Science

In principle, the H-D method is applicable to computer simulation modeling. The model is a hypothesized process and its predictions of the state variables (usually over time) are the test consequences. Modelers call the process of comparing predicted and observed test consequences "verification" or "validation," and a valid model is one whose predictions are within an a priori designated tolerance. However, when the tolerance characterizing the use of the H-D method in ecological studies that are not based upon modeling (e.g., testing salmon navigation hypotheses) is used for hypotheses composed of computer models, it is almost certain that no models can ever predict well enough to be declared valid. Herein lies a dilemma: modeling seems to hold promise for science because models can incorporate many features of the natural world. What are felt to be the essential features of a deer herd (Medin and Anderson 1979) or, indeed, an ecosystem (Innis 1978) can be combined in 1 large set of interrelated equations; however, the more equations and parameters a model contains, the more computation that is required, and the greater will be the propagation of errors in the parameters from model input to model output (Alonzo 1968).

The scathing criticism of computer simulation models delivered by Wildavsky (1973) and Berlinsky (1976) is predicated on the fundamental law that errors in input to calculations must necessarily increase as calculations are made, and they point out that models linking scores and perhaps hundreds of equations representing process functions, fueled by as many or more estimated parameters, many often from expert guesses and all with systematic and random error, cannot predict well enough to be declared valid, except by chance, according to the standards of experimental science. Looked at another way, if the testing of the threshold-of-security hypothesis, as outlined earlier, is expected to be marginally conclusive because a few uncontrollable boundary-condition assumptions may not be satisfied, how can a vastly more complex idea embodied in a computer simulation model be subject to a conclusive test?

Modelers recognize this problem, for most modeling studies are marked by either no attempt to validate or validation in which the acceptable tolerance between predicted and observed consequences is decided after the comparison has been made, leading to biased claims of validity usually phrased as "reasonable," "not reasonable," and "satisfactory." Any of these severs the chain of feedback necessary to make science a self-correcting process. Modeling as used in this way is the method of retroduction, not the H-D method. The model is an informed guess, a mixture of knowledge and error, about a process of nature. Running the model on a computer shows, by deduction, what the informed guess entails. Moreover, models are often tuned or calibrated, a process in which some of the model parameters are "fiddled with" to force better agreement between predicted and observed consequences (Innis 1978, Scavia and Robertson 1979). But this also severs the chain of feedback, for what is being validated is not the modeler's ability to hypothesize models that turn out to be valid, but rather his ability to fiddle.

Even when models are left unvalidated or fail to pass strict validation standards, modelers still believe that models are valuable for gaining insights and for sensitivity analysis, a process in which the changes in model outputs as a function of given changes in model parameters is gauged; the most critical parameters are then made the most important candidates for obtaining improved estimates. But, a model of doubtful validity can only give doubtful ideas, whether they are ideas called insight or ideas about parameter sensitivity.

## Computer Simulation Modeling as a Planning Tool

What is more often referred to now as modeling started out by being called systems analysis (Eberhardt 1975). Systems analysis is an organized method of thinking about planning; it follows the sequence: definition of a problem, identification of objectives, modeling, evaluation of alternative images of the future produced by subjecting the model(s) to different imagined controls, and implementation of results (Jeffers 1978). Because modeling is central to systems analysis, its name dominates. Thus, modeling's proper place is in the planning process for predicting alternative images of the future. As a planning tool, model calibration and sensitivity analyses are not only legitimate, but are demanded, for they can be expected to increase predictive capabilities.

Models in planning compete with traditional tools like regression analysis. Models often hold an advantage in that they can usually be made to predict the kind of information a wildlife manager needs for planning. For example, the deer population model of Medin and Anderson (1979), although an inaccurate predictor by the standards of normal science, does predict essential information a manager needs to know to set harvests, namely, the number of deer in the future with and without harvest. Regression analysis would have trouble answering this question. Modeling can usually be made to give a direct but inaccurate answer to a question, whereas other planning tools like regression analysis can usually be made to give an indirect but more accurate answer. The relative values of directness in answering a question, accuracy, and costs need to be addressed before deciding which tool to use. There will be a real situation in which each tool is best, so there is a need to learn and maintain all tools.

Modeling was never intended to function as a means to scientific knowledge. Its use in science is limited because it usually cannot predict to within established tolerances. However, as a comprehensive design and planning tool that can integrate scientific and common sense knowledge as well as theory, hunches, and expert opinion to forecast alternative future images, its continued use is assured. Planning is inaccurate by the standards of science, but that is no reason to abandon planning.

## Validity of Planning Tools

The wildlife science profession needs standards for validating the use of planning tools. Medin and Anderson (1979) suggest that computer simulation models be validated according to their utility to decision-makers. I think this is a correct concept, but is worthless until the profession operationally defines utility.

The problem of setting standards and holding planners accountable to these standards afflicts every profession. In business, where product life cycles are relatively short, satisfaction of public tastes forms the standard. In natural resource management, where planning outcomes often do not take hold until after the planners have retired or died, where the outcomes are more abstract, and where public tastes change over long periods of time, the situation calls for a different kind of standard.

What should it be? Very little has been written on the subject, but I think the ideas of Brewer and Shubik (1979) on what standards ought to be for computer simulations of mathematical war games form a good starting point. Their central concept is that models should receive professional evaluation by disinterested parties and by the eventual users throughout their model development phase, and modeling practitioners should spend more time defending their models than developing and using them. The principle is that if it is impossible to check on the outcomes of an implemented plan, then the utmost care should be taken to assure that planning models incorporate knowledge and exclude human error and biases. The difficulty lies not with the nature of the standard, which is common sense, but with holding practitioners accountable to it.

## CONCLUSION

Because the wildlife literature is taken as a role model for what wildlife science ought to be, and because it does not place the H-D method in a prominent role, widespread use of the H-D method is not guaranteed. I think the natural place to break this circle is in university education. As it now stands, education in the natural resource fields (almost everything I've said in this paper applies to the way all environmental sciences conduct themselves) does not provide training in scientific methods. Many, if not most, wildlife graduate students do not even understand the differences between induction and deduction.

Training is needed in all phases of science, and these principles need to be carried through in all wildlife courses. Students must be trained in the creative arts of asking the right questions, creating research hypotheses, using the scientific methods of induction and retroduction and the H-D method, designing efficient experiments (so as to avoid firing a cannon at a fly), and recycling the procedures so that the endless cycle of question and answer forms a unified whole. Students also must be trained in the ethics of science and planning; their teachers need to demonstrate these ethics in living form.

Wildlife science must try the H-D method. Without it the ability to detect errors in pronouncements of laws, the self-correcting feature science must have, is fatally lacking. All learning takes place in a feedback system in which ideas and reality interplay. The method of retroduction coupled with the H-D method is such a feedback system. Uncouple them and the ability to learn, to tell error from truth, is hindered, if not destroyed.

By themselves, scientific methods are impotent. Skills in using methods are the catalysts of potency. If, in a half century, the H-D method has been tried and shown to be impotent, then its judges must show that the cause was not the impotency in the skills and dedication of those who tried it.

I regard medical science and wildlife science as fields with equal potentials for achieving reliable knowledge. I think, however, that medicine has come closer to its potential, whereas wildlife science has lagged. I think medicine owes its success to the strict attention it pays to scientific method. Scores of books on the philosophy of clinical experiments have been published, yet I know of few comparable books in the natural resource sciences. Medical science obviously cares for and is committed to the quest for reliable knowledge. It is a good role model.

## LITERATURE CITED


ALONZO, W. 1968. The quality of data and the choice and design of predictive models. Pages 178–192 *in* G. C. Hemmens, ed. Urban development models. Highway Res. Board, Washington, D.C., Spec. Rep. 97. 266pp.

BAKER, J. J. W., AND G. E. ALLEN. 1968. Hypothesis, prediction, and implication in biology. Addison-Wesley Publ. Co., Reading, Mass. 143pp.

BERGERUD, A. T. 1974. Decline of caribou in North America following settlement. J. Wildl. Manage. 38:757–770.

BERLINSKY, D. 1976. On systems analysis. MIT Press, Cambridge, Mass. 186pp.

BOAG, D. A., AND V. LEWIN. 1980. Effectiveness of three waterfowl deterrents on natural and polluted ponds. J. Wildl. Manage. 44:145–154.

BOULDING, K. E. 1956. The image. Univ. Michigan Press, Ann Arbor. 175pp.

———. 1980. Science: our common heritage. Science 207:831–836.

BREWER, G. D., AND M. SHUBIK. 1979. The war game. Harvard Univ. Press, Cambridge, Mass. 379pp.

CAUGHLEY, G. 1980. [Book review of] The George Reserve deer herd. Science 207:1338–1339.

CHITTY, D. 1967. The natural selection of self-regulatory behaviour in animal populations. Proc. Ecol. Soc. Aust. 2:51–78.

EBERHARDT, L. L. 1970. Correlation, regression, and density dependence. Ecology 51:306–310.

———. 1975. Applied systems ecology: models, data, and statistical methods. Pages 43–55 *in* G. S. Innis, ed. New directions in the analysis of ecological systems, Part 1. The Society for Computer Simulation, La Jolla, Calif. 132pp.

EDWARDS, R. Y., AND C. D. FOWLE. 1955. The concept of carrying capacity. Trans. North Am. Wildl. Conf. 20:589–602.

EMLEN, J. M. 1973. Ecology: an evolutionary approach. Addison-Wesley Publ. Co., Reading, Mass. 493pp.

ERRINGTON, P. L. 1945. Some contributions of a fifteen-year local study of the northern bobwhite to a knowledge of population phenomena. Ecol. Monogr. 15:1–34.

FEINSTEIN, A. R. 1977. Clinical biostatistics. The C. V. Mosby Co., St. Louis, Mo., 468pp.

FELD, B. T., AND K. TSIPIS. 1979. Land-based intercontinental ballistic missiles. Sci. Am. 241(5):51–61.

FREUND, J.E. 1962. Mathematical statistics. Prentice-Hall, Inc., Englewood Cliffs, N.J. 390pp.

GILES, R. H. 1978. Wildlife management. W. H. Freeman & Co., San Francisco, Calif. 416pp.

HANSON, N. R. 1965. Patterns of discovery. Cambridge Univ. Press, Cambridge, U.K. 241pp.

HARVEY, D. 1969. Explanation in geography. Edward Arnold, London, U.K. 521pp.

HASLER, A. D. 1966. Underwater guideposts. Univ. Wisconsin Press, Madison. 155pp.

———, A. T. SCHOLZ, AND R. M. HORRALL. 1978. Olfactory imprinting and homing in salmon. Am. Sci. 66:347–355.

HAYNE, D. W. 1978. Experimental designs and statistical analyses in small mammal population studies. Pages 3–10 *in* D. P. Snyder, ed. Populations of small mammals under natural conditions. Univ. Pittsburgh Press, Pittsburgh, Pa. 237pp.

INNIS, G. S. 1978. Grassland simulation model. Springer-Verlag, New York, N.Y. 287pp.

JEFFERS, J. N. R. 1978. An introduction to systems analysis: with ecological applications. University Park Press, Baltimore, Md. 198pp.

KREBS, C. J. 1979. [Book review of] Populations of small mammals under natural conditions. Science 203:350–351.

LAUCKHART, J. B. 1955. Is the hen pheasant a sacred cow? Trans. North Am. Wildl. Conf. 20:323–336.

———, AND J. W. MCKEAN. 1956. Chinese pheasants in the Northwest. Pages 43–89 *in* D. L. Allen, ed. Pheasants in North America. The Stackpole Co., Harrisburg, Pa., and The Wildl. Manage. Inst., Washington, D.C.

LEOPOLD, A. 1933. Game management. Charles Scribner's Sons, New York, N.Y. 481pp.

MEDAWAR, P. B. 1969. Induction and intuition in scientific thought. American Philos. Soc., Philadelphia, Pa. 62pp.

MEDIN, D. E., AND A. E. ANDERSON. 1979. Modeling the dynamics of a Colorado mule deer population. Wildl. Monogr. 68. 77pp.

NELSON, N. M., M. W. ENKIN, S. SAIGAL, K. J. BENNETT, R. MILNER, AND D. L. SACKETT. 1980. A randomized clinical trial of the Leboyer approach to childbirth. New Engl. J. Med. 302:655–660.

PLATT, J. R. 1964. Strong inference. Science 146:347–353.

POPPER, K. R. 1962. Conjectures and refutations. Basic Books, New York, N.Y. 412pp.

RACHELSON, S. 1977. A question of balance: a wholistic view of scientific inquiry. Sci. Educ. 61:109–117.

SCAVIA, D., AND A. ROBERTSON. 1979. Perspectives on lake ecosystem modeling. Ann Arbor Science Publ. Inc., Ann Arbor, Mich. 326pp.

TAYLOR, M. W., C. W. WOLFE, AND W. L. BAXTER. 1978. Land-use change and ring-necked pheasants in Nebraska. Wildl. Soc. Bull. 6:226–230.

VERME, L. J., AND J. J. OZOGA. 1980. Influence of protein-energy intake on deer fawns in autumn. J. Wild. Manage. 44:305–314.

WAGNER, F. H. 1969. Ecosystem concepts in fish and game management. Pages 259–307 *in* G. M. Van Dyne, ed. The ecosystem concept in natural resource management. Academic Press, New York, N.Y.

———. 1978. Some concepts in the management and control of small mammal populations. Pages 192–202 *in* D. P. Snyder, ed. Populations of small mammals under natural conditions. Univ. Pittsburgh Press, Pittsburg, Pa.

———, AND A. W. STOKES. 1968. Indices to overwinter survival and productivity with implications for population regulation in pheasants. J. Wild. Manage. 32:33–36.

———, C. D. BESADNY, AND C. KABAT. 1965. Population ecology and management of Wisconsin pheasants. Wis. Conserv. Dep. Tech. Bull. 34. 168pp.

WILDAVSKY, A. 1973. [Book review of] Politicians, bureaucrats, and the consultant. Science 182:1335–1338.

*Received 27 March 1974.*
*Accepted 10 August 1980.*