A RAND NOTE

Six (or So) Things You Can Do with a Bad Model

James S. Hodges

RAND

The research described in this report was supported by RAND as part of its program of public service.

This Note contains an offprint of RAND research originally published in a journal or book. The text is reproduced here, with permission of the original publisher.

Published 1991 by RAND
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138

A RAND NOTE

N-3381-RC

Six (or So) Things You Can Do with a Bad Model

James S. Hodges

RAND

SIX (OR SO) THINGS YOU CAN DO WITH A BAD MODEL

JAMES S. HODGES
RAND, Santa Monica, California
(Received September 1990; revision received March 1991; accepted March 1991)

Many models used in policy or systems analysis either cannot be validated in any fully adequate sense, such as by comparing them with actual data, or could adequately be validated but have not been. For example, in the area of combat analysis, the central models are arguably almost entirely unvalidated and most will never be susceptible to adequate validation. Nevertheless, such models are often used and can be used fruitfully, even though we have no theory for how to use them or how to interpret and place value on the results they produce. This paper takes a step toward providing such a theory by focusing on the logic that should govern the use of inadequately validated models and the costs and benefits of using them. To this end, it identifies and evaluates six legitimate uses to which such models can be put.

Often a policy or a systems analyst has data on inputs to a system but not on its outputs. (By the term data I mean a statistician's measurements of real things rather than a computer scientist's strings of characters or an operations researcher's variously obtained parameter values.) If the analyst has an empirically valid model of the system, this lack does not pose a major problem: A model that has been shown to turn inputs into outputs faithful to reality can often be used confidently in situations in which the analyst only has inputs and the model.

Often, though, a policy or systems analyst is stuck with a bad model, that is, one that appeals to the analyst as adequately realistic but which is either: 1) contradicted by some data or is grossly implausible in some aspect it purports to represent, or 2) conjectural, that is, neither supported nor contradicted by data, either because data do not exist or because they are equivocal. For example, most military combat models are bad models in at least one of these senses. A model may have component parts that are not bad, but if, taken as a whole, it meets one of these criteria, it is a bad model. (The term bad model may seem polemical. It is intended to be.)

Just because a model is bad, however, does not mean it is useless. Indeed, this paper is about legitimate uses of bad models, in the following sense. An argument's path from premises to conclusions must be logically clear and correct. A bad model can be used to construct correct paths from premises to conclusions, but because its relations to reality are questionable, it can only do so in a few ways—at least, ways that permit useful conclusions with respect to reality. The purpose of this paper is to catalog and discuss logically clear and correct ways that bad models can do that. Thus, its object is to take a step toward a theory of bad models, founded on explicit treatment of the logic of their use.

Analysts can clearly use such a theory, and so can consumers of analyses. Analysts can use it to force themselves to be clear about the actual purposes of their models, thus avoiding the pitfall of having a model's contribution turn out to be less than is apparent and overpriced for what it is. Consumers of analysis can use such a theory similarly.

A bad model can be used legitimately—that is, with clean logic—in six ways (in rough order of increasing complexity, subtlety, and susceptibility to facile rationalization):

1. as a bookkeeping device, for example, by condensing masses of data, and by providing a means or incentive to improve data quality;
2. as part of an automatic management system whose

Subject classifications: Professional: comments on. Military: warfare models.

efficacy is not evaluated by using the model as if it were a true representation;
3. as a vehicle for *a fortiori* arguments;
4. as an aid to thinking and hypothesizing, for example, as a hypothetical statement as the basis for purely intellectual explorations, as a stimulus to intuition in applied research or in training, or as a decision aid in operating organizations;
5. as an aid in selling an idea of which the model is but an illustration;
6. as a training aid, to induce particular behavior.

Once the logic of the use of bad models is explicit, obscure issues can be clarified. This paper discusses two such issues: model validation and the costs and benefits of model use. After some general points on these issues in this introductory section, Section 1 discusses the six legitimate uses and takes up the validation considerations specific to each. Section 2 covers cost-benefit issues, and Section 3 offers some conclusions. Throughout the paper the primary example is military combat models.

Validation or (Preferably) Evaluation

In this paper, the term verification refers to "the process by which the analyst assures himself and others that the actual model that has been constructed is indeed the one he intended to build" (Miser and Quade 1988, p. 528). This process includes, among other things, assessing the model's internal consistency and the correctness of its corresponding computer code (Miser and Quade 1988, pp. 530–534).

On the other hand, the term validation refers to "the process by which the analyst assures himself and others that a model is a representation of the phenomena being modeled that is *adequate for the purposes* of the study of which it is a part" (Miser and Quade 1988, p. 529, emphasis as in the original). Many treatments of validation begin with a two-part definition like this, but then emphasize the first of the two parts, the representation, and downplay the second, the purpose. For example, Thomas's lucid paper repeatedly refers to validation as "testing the 'agreement' of the model with 'reality'" (Thomas 1989, p. 260, quotes as in the original). Beginning with this premise, Thomas concludes that validation cannot even be defined, much less executed, and is left with the prescription that we must strive to validate even though it is impossible.

This unhappy conclusion is not inevitable; it follows from the choice to focus on the first part of Miser and Quade's two-part definition. The key to validation is the second part of the definition, that the model be *adequate for the purpose of the work*. Miser and Quade (1988) downplay this somewhat by emphasizing uses of models for predictions (*loc. cit.* p. 528). For predictive purposes, validation is fruitfully understood as a process by which a model's predictions are subjected to a series of tests, with confidence in the model growing with the amount and variety of tests that its predictions survive. (See Miser 1988, pp. 469–489, and Miser and Quade 1988, pp. 534–539, which draw heavily on the work of Ravetz.) For nonpredictive purposes, however, this approach does not necessarily illuminate the problem of validation.

This paper emphasizes the role of purpose by restricting consideration to models that are either unvalidated—that is, that have accrued little confidence in Miser and Quade's terms— or *in*validated, in the sense of having decisively failed some predictive tests. Such models are not necessarily useless; in fact, when each of the six possible uses or purposes is considered, it becomes clear that the adequacy of the model for the purpose does not necessarily involve any testing of the agreement of the model with reality. Indeed, in the last of the six uses, substantive deviations from reality are purposefully introduced.

The approach taken here could thus help to cut the Gordian knot of validation: When predicting, the criterion is the adequacy of the predictions, as discussed by Miser and Quade; for other uses, the standard depends on the use, as detailed in the sections to follow. Because the term validation has traditionally been reserved for models in predictive uses, I prefer to use the term evaluation for nonpredictive purposes, so as not to dilute the traditional force of "validation."

Some readers have argued that the criticism implied by the term "bad models" is undeserved because they can be used appropriately in some cases. I agree, partly: This paper is, indeed, about the logic of their appropriate use. If the logic works, the use is appropriate; if it fails, the use in inappropriate. (Cost effectiveness is a separate issue.) As for the pejorative connotation of the term bad model, perhaps we should admit that many useful models would be embarrassments to scientists, from whom we got the idea of a model, but whose job is to improve the match between models and reality.

Cost-Benefit Considerations

The use of models consumes resources and sometimes yields benefits. Once stated, this is obvious, but it tends to be obscured by an intellectual tradition that ignores the costs of conducting analyses and assumes that they nearly always yield significant benefits. The costs of using models in making analyses include dollar costs and subtler opportunity costs. As for benefits, the interests of analysts diverge from those of their clients. Clients want information they can trust to help them develop perceptions of their problem situation and, subsequently, of policies and decisions. Analysts want to deliver timely and cost-effective information related to their client's problem, to sell the current analytical product, to secure future work, and sometimes to grind personal axes, such as using a pet method. This list is not cynical: Even Mother Theresa has to stay in business, and a job done right but sold poorly might as well not be done. Indeed, if the work is good, a client has an interest in having it sold effectively, although if the selling methods are too expensive the client might benefit by learning how to be sold more economically. Further discussion of costs and benefits is deferred to Section 2.

Things Not Considered Here

A modeler can face different degrees of uncertainty:

- no empirical evidence at all, so that modeling is theology;
- some empirical evidence, although few, if any, specifics are possible without assumptions of unfounded precision, as in combat modeling;
- relationships are empirically sound, but parameter values are uncertain;
- relationships and parameter values are empirically sound.

This paper is about situations that exhibit the first two of these degrees of uncertainty. The fourth is logically the simplest and most desirable, but I have never seen an example of it in policy or systems analysis. The third is rare but does occur, for example, in formulas relating an aircraft's fuel capacity and design to its range and payload. Because such models are rare and allow more confident uses than do bad models, this paper will not consider them. Many modeling papers suggest that their writers believe they face uncertainty about parameters only, when a little questioning makes it clear that analyses intended to shed light on most problems are more difficult than this.

1. SIX (OR SO) LEGITIMATE USES OF BAD MODELS

This section discusses the six legitimate uses of bad models listed in the Introduction, and the form of evaluation appropriate to each.

As a Bookkeeping Device

By condensing masses of data. Some models digest prodigious volumes of inputs and produce neat summaries. The summaries can be useful quite apart from anything else the models do. For example, a RAND-developed collection of programs called DRIVE—which schedules component repair and distribution for Air Force repair depots—produces handy displays of the location of broken and usable parts throughout the Air Force's logistics system. Even if repair managers ignored DRIVE's scheduling features, the displays would be an advance over previous tools.

A model can serve a different bookkeeping function by helping an analyst organize data. A colleague who uses strategic exchange models has described them as useful in this way. This user tried to extend the notion of bookkeeping to his model's outputs as well, on the grounds that strategic exchange models are essentially bean-counting exercises. But translation of input data to outputs, via the bean-counting routines, requires assertions about the nature of strategic combat that are mostly without empirical support. In this case, then, the notion of bookkeeping applies only to the input data.

By providing a means or incentive to improve data quality. Analysts and managers are often inconvenienced by low quality data. Sometimes quality is low because data collectors have no incentive to record the data carefully. But a model that data collectors care about, and which needs good data, can create the right incentives. For example, Air Force base commanders and maintenance officers are becoming aware of the potential usefulness of the DRIVE model within theaters. Quality problems have plagued the data used by DRIVE, but interest in DRIVE has stimulated interest in the data it uses. While the original stimulus may have been provided by DRIVE, unrelated uses have since been found for the improved data. (This use and the example were contributed by Warren Walker.)

Evaluation. A bad model used to condense masses of data is evaluated by ensuring that it reads the right input numbers accurately and then summarizes and displays them correctly. This is identical to verification, as that term was defined in the Introduction. A bad model used to induce higher data quality is evaluated by comparing the accuracy of the data before and after the use of the model.

As Part of an Automatic Management System, Whose Efficacy Is Not Evaluated by Using the Model as if It Were a True Representation

Some models can be viewed reductively as algorithms that turn input numbers into output numbers. As such, a model can be inserted into a management system in which the outputs drive more or less automatic functions. For example, Kalman-filter and other time-series models are used to process data from sensors in freeway road surfaces, and in turn to run metered on-ramps. There is little reason to take the Kalman filter model at face value as a representation of traffic flow, but its performance can be judged easily enough in this situation.

In a like manner, the DRIVE routines just mentioned use a model of part failures to make short-term predictions of failures. This model implicitly assumes that part failures are more predictable than they actually are. (Decades of data consistently contradict each element of DRIVE's part failure model. For documentation, see Hodges 1989.) Nonetheless, a prototype of DRIVE has been implemented at the Ogden Air Logistics Center and early returns suggest that it schedules repairs in the advertised manner. The falsity of the part failure model, while unappealing, is irrelevant as long as the performance of the overall system can be evaluated independently of the model.

A management system driven by a bad model must not be tested by using the model as if it were true. By presumption, the model is a deficient picture of reality, and it presents the management system with the easiest possible test because it, unlike the cruel world, satisfies the system's assumptions. But a bad model can be used as a vehicle for *a fortiori* arguments in an evaluation of a system of which it is a part (as argued in the next subsection).

Evaluation. A bad model used as part of a management system is evaluated by measuring the efficacy of the management system. If it works, it works, and that is all that matters. Determining whether it works is not without subtlety. The fundamental problem—assuming measurement is done right—is to assure that test conditions adequately approximate those of actual use.

Contrary impressions notwithstanding, this is not a defense of black boxes. It is a defense of simple, dumb-looking, transparent boxes that do the job.

As a Vehicle for *a fortiori* Arguments

An *a fortiori* argument can work like this: If condition X were true, then policy A would be preferable to the other candidates. But the actual situation deviates from X in ways that favor A even more. Thus *a fortiori*, A is preferable.

A bad model may be used in an *a fortiori* argument. As noted, models of part failures used in Air Force spare parts analyses presume that part failures are more predictable than decades of research show them to be. Nonetheless, these bad models can be used to argue in support of supply systems that are flexible or responsive, as distinct from systems that rely on accurate predictions of part failures. If a responsive system is superior under conditions that are more predictable (i.e., less variable) than actual conditions—as standard models are—then *a fortiori* it will be superior under actual conditions. An *a fortiori* argument like this is implicit in analyses supporting CLOUT, a collection of RAND initiatives for improving the Air Force's spares and repair management system. (CLOUT is an acronym for Coupling Logistics to Operations to meet Uncertainty and the Threat; see Cohen, Abell and Lippiatt 1991.)

For another example, a RAND colleague who does Army research on new types of weapon systems has defended the JANUS combat simulation model on the grounds that it "limits the bullshit" of advocates of new technologies. This is an *a fortiori* argument: Actual combat would tax exotic systems more than JANUS does, JANUS finds exotic system A to be wanting, therefore, JANUS has "limited the bullshit" by rejecting exotic system A. (This observation is due to Dick Salter.) This argument, if correct, can be used to reject exotic systems but not to declare them worth buying.

Evaluation. An *a fortiori* argument has three parts: that condition X implies policy A is preferable, that X represents a boundary on the actual situation, and that reality's deviations from X favor A. Evaluation of a bad model for use in an *a fortiori* argument depends on how the bad model is used in each of the three parts. This discussion will cover evaluation for the two examples given above, because I know of no logically distinct *a fortiori* arguments using bad models. If others exist, they might require different forms of evaluation.

The first part of the argument, "X implies A is preferable," takes X as true and draws an implication from it. In the example, X is "the bad model accurately represents the actual situation." The burden of this part of the argument is drawing the implication correctly, and evaluation, as distinct from verification, plays no role. Indeed, the argument presumes that X is false in a specific way, so there is no reason to validate X in the usual sense. (In formal logic, implications are statements of the form "if A, then B," which impose relations between the truth values of statements A and B. The obvious relation is that if A is true, B is also, but if A is false, B can be true or false. The first part of the *a fortiori* argument uses the property that A can be false and B true without creating a contradiction.)

The second part of the argument, "X represents a boundary on the actual situation," does require that an assertion in the model be related to a fact in the world. In the JANUS model, a notional weapon system works as advertised, is not subject to Murphy's law, and so on. This makes JANUS a more benign environment than actual combat, so the assumption that JANUS accurately represents the situation facing an exotic weapon forms a boundary on the actual situation.

The third part of the *a fortiori* argument, "reality's deviations from X favor A," is likewise subject to evaluation. If deviations from X "go in one direction" in a sense meaningful in context, they must be shown to favor A. It is tempting to try to evaluate this part of the argument by using the bad model itself, as in: if we push parameter values away from the values given by X in the direction we know to be true, the model produces outcomes more favorable to A. This is fine as long as the model is bad only because its parameter values are unknown. It is not acceptable if the model is bad in some other way that casts doubt on the information value gained by changing the parameter values specified by X.

In both *a fortiori* arguments used as examples here, the third part of the argument is evaluated by an appeal to common sense, or to folk wisdom accumulated from related but usually simpler problems. For example, when the variability of part failures is increased, we believe that a responsive supply system does better than an unresponsive system because an unresponsive system cannot adapt to the dislocations caused by the variability and a responsive system can. As long as the common sense behind such arguments is common sense about the world *and not about the model*, appealing to it is acceptable, although not squeaky clean.

The more complicated a model is, the harder it is to evaluate the second and third parts of an *a fortiori* argument. As a model grows in complexity, its real structure becomes more obscure, because interactions become more difficult to grasp and the computer code is more prone to undetected errors. These effects are familiar to software engineers but apparently not to policy modelers. *A fortiori* arguments are also likely to become more difficult as more parameters are added because, generally, more parameters are unknown and must be bounded in the "right" direction.

As an Aid to Thinking and Hypothesizing

A bad model is a combination of assertions, some factual, others conjectural, and others plainly false but convenient. What, then, are we to make of the ubiquitous claim that bad models provide insight? *Webster's Ninth New Collegiate Dictionary* defines *insight* as: "The power or act of seeing into a situation: Penetration" (p. 626), and elaborates this with:

> *Discernment* stresses accuracy (as in reading character or motives or appreciating art); . . . *insight* suggests depth of discernment coupled with understanding sympathy (p. 360).

By definition, a bad model does not give power to see—accurately, deeply, or at all—into the actual situation, but only into the assertions embodied in the model. Thus, if the use of a bad model provides insight, it does so not by revealing truth about the world but by revealing its own assumptions and thereby causing its user to go learn something about the world. Three such instances are sufficiently distinct to deserve separate discussion:

- as a hypothetical statement as the basis for purely intellectual exploration;
- as a stimulus to intuition in applied research or in training;
- as a decision aid in operational settings.

As a hypothetical statement as the basis for purely intellectual explorations. A model is a straw man, a group of null hypotheses. It provides a collection of questions that need to be answered. To the extent that it organizes what is believed and known, it can structure data, debate, and teach although it also constrains those activities.

This argument has been used by a RAND researcher to defend bad models generally and combat models in particular. As presented, it is unassailable as a defense of *building* bad models. But if a bad model, once built, is used to draw conclusions or advice about something in the work, then it is no longer being used merely to organize beliefs and knowledge, and its use must be justified by some other argument. I have yet to see a policy model built without some intent to influence decisions. Thus, although it is legitimate to use a bad policy model to summarize belief and knowledge, this use is probably not important in practice.

As a stimulus to intuition in applied research or in training. Sometimes it can be useful to draw implications from the assertions in a bad model. Many are intuitively obvious, and nothing is learned. Some are not obvious, or rather, they conflict with a prior belief. Of these, most are errors in data or computer code, or artifacts of a specific assumption representing a vague belief. These implications disappear on examination. Some implications, however, do not disappear, and are striking. At this point, the model's user has only learned something about the assumptions in the model: it is arithmetic, not science. If the user is then moved to go learn something about the world, the model may be said to have provided an insight by poking its user to go look at something out in the world. But it cannot be overemphasized that the model only tells the user about its own assumptions and *not necessarily about the world*: A bad model can suggest, but it cannot reveal truth. That must be found elsewhere.

A program manager at RAND defends a recent well-briefed use of the TAC-SAGE combat model in the foregoing terms. He argues that the conclusions suggested by the model were either justified independently of the model or were not reported. The model, then, was used to suggest implications of assumptions, but not to justify them as properties of actual combat or weapon systems. Justifications were by appeals to common sense, by deductions from simpler models, or by other appeals to evidence external to TAC-SAGE.

A variant on the above is that an idealized model, like the perfectly competitive market, might suggest ideas or policy alternatives that would otherwise be hard to come by. As before, the model suggests, but truth must be found elsewhere.

This stricture—that a bad model can only suggest—is stronger than it may appear. Bad models produce numbers, and thus present an unbearable temptation to use those numbers as if they do more than suggest. They cannot. If a model is bad as defined here, and the specific numbers it produces cannot be buttressed by some other arguments, then *the numbers have no meaning except as illustration of the consequences that flow from the model's assumptions.*

The aforementioned program manager argues that some decisions require specific numbers, and that production of numbers seems to require models. His example was the ratio of U.S. to Soviet reductions in asymmetric arms reduction treaties. He argues that if a number would not be accepted without a model's imprimatur, no harm is done by using a model to produce it so long as the analysts involved are fastidious enough in satisfying themselves. In this case, illustration veers into obfuscation, namely of the basis of the specific number. It may be necessary for sales purposes to certify that a number was produced by a computer, but in the words of Orwell (in the essay "Writers and Leviathan"), we should not conclude that if a thing is necessary it is also right.

The logic of the foregoing discussion applies almost without change to the use of bad models for training. A retired Air Force pilot at RAND says that the JANUS model gave him his first opportunity to attack a column of tanks. The model permits low-risk, relatively cheap trial-and-error learning. The question is whether the pilot learns about an actual attack on real tanks or only about some modeler's rendition of it. Again, the model can only suggest.

As a decision aid in operational settings. A bad model can serve as a decision aid to someone in, for example, a military staff position. A RAND project developed a spreadsheet model intended to help the Air Force's theater staffs consider wartime redistributions of logistics assets. The model's developers make no pretense that it is a good model; it is not supposed to think for staff planners, only to help. It does this in much the same manner as above; it suggests redistribution actions, makes a rough cut at the required transportation load, and presents these and other things to the planner. The model might suggest things the planner would never think of because of its speed and thoroughness. It might also suggest things the planner would never *want* to think of because it is a bad model. Evaluations of the model's suggestions must be done outside the model because the planner knows things that it cannot know, and this is as it must be.

Evaluation. If a bad model is set forth merely as a hypothetical statement as the basis for intellectual exploration, I cannot cavil, but if it slips into use as a statement of belief and knowledge about reality, then it must be evaluated by ensuring that it satisfies the logic of this more ambitious use. This case will be discussed because it is relevant to the cases of stimulus to intuition and decisions aids.

For many models, knowledge (in the sense of science; see Miser 1988, pp. 469–489) is scarce, so evaluation consists mainly of ensuring that the computer code represents beliefs accurately. This reduces to verification in many cases, although the representation of beliefs can present two difficulties. First, many beliefs are vague, as in "forces move slower in difficult terrain." To be represented in a computer program, this belief must be made precise by grading difficulty of terrain and movement of forces, and specifying a relationship between them. (Apparent alternatives like probability or fuzzy sets just substitute one precise specification for another.) If a model is to be built, then vague beliefs must be given specific forms that are essentially arbitrary even if they do not contradict the vague beliefs they represent. In such cases, evaluation consists of ensuring that the specifics of the model do not contradict the vague beliefs that motivate them.

One might hope for a standard of consistency instead of a lack of inconsistency, but, as a practical matter, no one can make every consistency check, so a stingent lack of inconsistency is the most one can ask for. Even this need not be simple. For example, recent research at

RAND (Dewar, Gillogly and Juncosa 1991) shows that simple attrition combat models that use thresholds (e.g., Blue withdraws if the ratio of Red's forces to Blue's becomes worse than n:1) display grossly counterintuitive outcomes, in particular, extreme sensitivity to minute changes in some inputs. Each individual belief/assertion of such a model may be valid in the sense of the last paragraph, but the ensemble fails the test of "do not contradict the motivating vague belief." Combat modelers may decide that combat is chaotic (in the technical sense) so that these preliminary findings will cease to be disturbing. But that would require either establishing a fact about the world—that combat is chaotic—or a change of beliefs; that is, the modelers must change the standard against which the model is evaluated.

The second difficulty with evaluating a model used to summarize beliefs is that a researcher may wish to summarize only some of his beliefs, because otherwise he would need all the computing power in the world. How can a model omit phenomena and remain true to the researcher's beliefs? Suppose, for example, that a researcher wished to assemble his beliefs about ground combat without going into detail about the effects of air power. If the model treated ground warfare as if air power did not exist, it would surely contradict the researcher's beliefs. Thus, the model must represent the effects of air power implicitly, by (for example) using parameter values different from those that would be used in a model that treated air power explicitly. (This example is due to Bob Levine.) In general, the effects of an omitted phenomenon like air power will not be simple, so its implicit representation must involve averaging over some set of cases or using the effect that obtains a most likely case. Evaluation then becomes: specify the cases averaged over or the most likely case, and judge whether the model represents belief for these cases. It is hard to imagine an instance in which this will not put the researcher in the position of evaluating a vague belief, as discussed previously.

Sometimes the proprietors of a bad model claim that parts of it are facts, not just beliefs. Evaluation then amounts to determining if facts support the claims, and disciplines like statistics have tools for this task. The difficulty of using statistical tools will vary depending on the problem. In the case of atmospheric models, some kinds of data are ample, ably collected, and readily interpretable. For combat models, however, most of the data come from earlier wars and are spotty, badly collected, and of debatable relevance. This does not imply that such data should be ignored, but they are sometimes used naively, and that is a great danger.

Many of the foregoing considerations apply to bad models used in applied research, in training, or as decisions aids. In these cases, however, the objective is to stimulate thought, and greater inaccuracy is tolerable. Indeed, because the model's suggestions must be evaluated by means external to it, the criterion for judging the model should be efficiency—the largest number of good ideas for a given amount of resources—with superficial realism having no intrinsic value. This will be pursued in Section 2 under cost-benefit considerations.

As an Aid in Selling an Idea of Which the Model is But an Illustration

Architects build scale models of new projects so that developers, financiers, and city officials can see how the projects will look. A scale model is a bad model in the sense used here: It is grossly discrepant with reality, if only because it is far too small for anyone to live in. Nonetheless, it can do a good job of selling the idea—the project—of which it is but an illustration, by conveying aspects of the idea concretely.

Mathematical models can serve this same function. A project at RAND is conceiving a spreadsheet-like model of the Navy's aviation logistics system that would allow its user to trade off expenditures for physical distribution, say, against expenditures on stocks of spare parts. In our armed forces, this idea is exotic and unfamiliar, and the first thing we will need to do is sell it. Even a crude version of such a model would be a powerful marketing tool, and once the idea is sold, it will be up to the services to take care of the details.

The last paragraph illustrates the danger of this use of a bad model: It almost begs to be used disingenuously. A model may be fine as a descriptive tool ("here are some things in your logistics system that you are trading off whether you know it or not") but poor as a predictive tool ("here is how much money you will save if you make the tradeoff this other way"). The requirements of the two uses are very different but they are not always distinguished. If we succeed in building our Navy model, it will most likely be a bad model. If we sell the idea of tradeoffs, and we would like it to be a predictive tool, but we will need some other logic to justify that use. That logic has not been devised because the model does not exist, and it may not be possible to devise adequate logic. If not, then our model can only be used to sell the idea of tradeoffs, not to make them.

Evaluation. A bad model used to sell an idea need only represent the idea and display benefits. Evaluation consists of ensuring that the model does both things, and this is all that can be asked of a model in this role. The analyst-cum-salesman is not off the hook: He had better believe that the benefits are real, and must accompany presentations with appropriate caveats. In the case of the architect's model, these warnings would sound like, "this model is only intended to show how the project looks; do not attempt to live in it." It sounds silly, but it is the correct form for caveats that must accompany bad models used to sell ideas.

As a Training Aid, To Induce Particular Behavior

Railroad engineers are trained partly in railroad engine simulators. These simulators are usually realistic, but not always. For example, if presented as a moving picture, the movement of telegraph poles past the side windows of the engine can "strobe" at certain simulated speeds, and this is very distracting for trainees. Thus, instead of simulating the apparent movement of telegraph poles, some simulators run past the side windows a black-and-white zebra pattern that is undistracting at any speed. In this case, a deliberately unrealistic model is used because the avoidance of distraction is more important than realism in the engineer's peripheral vision. (This use and example were contributed by Jim Dewar.)

In the U.S. Army, brigades train at the National Training Center (NTC), a 1,000 square mile piece of desert with a home team (the OPFOR) trained to fight Soviet tactics. The OPFOR is extremely proficient and has other advantages, such as familiarity with the terrain. This deliberately unrealistic aspect of the NTC is maintained because the trainers do not want Blue units to make mistakes without paying for them.

While this use is related to the third use, the *a fortiori* argument, it is distinct and may be antagonistic to that use. The third use is analytic—drawing analytical conclusions from an unvalidated model—while the present use has a training purpose. The ability to use the NTC analytically is diminished by the OPFOR's skill, because some apparent outcomes can plausibly be caused by the OPFOR's unrealistic advantage, and not by anything inherent in U.S. Army doctrine, tactics, or equipment.

Evaluation. The use of a deliberately unrealistic model to induce particular behavior is evaluated by determining whether it induces the desired behavior. The mechanics of this are familiar and will not be discussed. Note that not only is realism unimportant, it is deliberately sacrificed.

2. COSTS AND BENEFITS OF USES OF BAD MODELS

After some general comments about costs and benefits, the six legitimate uses of bad models will be discussed in turn.

General Considerations

Dollar costs of building and using models are fairly straightforward, including capital costs for equipment and software, operating costs, and maintenance costs. It should not be difficult to assess these costs for almost any model's use. Opportunity costs are more elusive. For example, an organization's talent mix is affected by dependence on computer-intensive modeling. Elaborate models require that the talent mix shift away from general purpose analysts and toward special purpose technicians. In the short run, the organization incurs a cost in that it has fewer analysts to call on for quick-delivery analyses; in the long run, clients' problems look more like nails waiting to be struck by the (big-investment) modeling hammer.

An analyst's interests can be analyzed as delivering a timely and cost effective solution to the client's problem, selling the current product, securing future work, and grinding personal axes, e.g., advancing a pet method. The contribution of a specific model to a specific job depends on what the job is, so it will be taken up below in considering the six legitimate uses. An analyst's cost-benefit calculation can loosely be understood as:

$$\text{payoff} = \text{benefits if sold} \times \text{probability of sale}$$
$$- \text{cost of sales technique}.$$

Some sales techniques are better than others with some audiences; thus, sales can be aided or impeded by the complexity of the model. The cost of the sales technique depends on the model's complexity and on gewgaws like computer graphics. A model helps to secure future work by the appearance of doing the job right and by selling it well, with the possible exception that a model billed as comprehensive may attract more work than one billed as custom-made for a given job. Finally, grinding personal axes is a waste of the client's money and an indulgence by the analyst. If the analyst is a calculating sort, he will calculate how much he can waste and still get away with it. Further consideration of this is outside the scope of this paper.

Cost-Benefit Considerations Specific to the Legitmate Uses of a Bad Model

Costs and benefits can be viewed in two ways: in terms of total costs and benefits, or on the margin. For the most part, the following discusses total costs and benefits, but for some uses marginal costs and benefits are discussed.

1. **As a bookkeeping device**. A complex model is an expensive way to get neat data summaries or to improve

data quality, but if the model exists for some other reason and is maintained, the bookkeeping function comes almost for free. It does not make sense to buy a radio, break it, and use it as a doorstop, but if I need a doorstop and all I have is a broken radio, it will hold my door.

2. **In an automatic management system**. This use of a bad model presents the easiest problem of measuring benefits. If a new model is used, say, to manage an inventory, savings can be measured straightforwardly. This points to an important truth: cost effectiveness and appearance have no necessary connection. It may be that improving the model would improve the system's performance, but it need not, and it would not necessarily be cost effective to use a better model even if it did improve the system's performance. It might be cost effective to use a worse model, if it were a lot cheaper to run and at worst only a bit less effective.

A model's contribution to selling itself is also straightforward for this use. If the client is alert and cares about obtaining benefits, the model sells if it yields real benefits. If the client is not alert or is not after real benefits, the model sells if it has plenty of knobs and flashing lights.

3. **As a vehicle for an *a fortiori* argument**. *A fortiori* arguments are usually not available at any cost. If one is available and correct, how much is it worth? This problem is very hard, and will be avoided here by the assumption that an analyst has chosen to make an *a fortiori* argument, and is considering more and less costly ways to do it. Then cost-benefit considerations can be examined on the margin. Start with a given model and thus cost. The model can be made simpler and cheaper, or more complex and expensive. Whichever direction is chosen, the *a fortiori* argument, which works for the given model, can fail because of the change or remain available but be more or less expensive, or a different and stronger argument might become available. It is difficult to generalize beyond recalling that greater complexity tends to make *a fortiori* arguments more difficult, so to this extent consideration on the margin will favor simplicity.

When it comes time to sell an *a fortiori* argument, transparency works. This will depend on the client's sophistication, but *a fortiori* arguments are a bit subtle and it is a safe bet that the more transparent, the better. Thus appearance—which usually means complexity and cost—need not increase the probability of sale and may decrease it. For the above reasons, one may hypothesize that for *a fortiori* arguments small is beautiful.

4. **As an aid to thinking and hypothesizing**. *As a hypothetical statement as the basis for purely intellectual explorations*. This use is most difficult to which to assign value. How does one value a scientific theory? Because this case is difficult and is not important for this paper, it will not be discussed further.

As a decision aid in operational settings. Measurement of benefits is relatively straightforward in this case, the issues are much like those discussed under Use 2, in an automatic management system. The difference is that for the present use it involves assessing the gain in efficiency of the model's user, i.e., how much is he aided by the decision aid? The benefits can be negative if the decision aid slows him down or misdirects him. Sales considerations are the same here as they were for Use 2: The sales payoff of different kinds of models depends on how alert the client is and how much he cares about benefits.

As a stimulus to intuition in training. For some tasks, it is uncomplicated to measure performance before and after training by a bad model and by some other method, and to compare the two. But many tasks are not so tractable. In the U.S. Army, for example, corps and division staffs train at the Warrior Preparation Center and the Battle Command Training Program by fighting "wars" on large simulation models, and brigades train at the National Training Center. The NTC alone has cost the Army billions of dollars, and while the majority opinion is that it is the finest training to be had anywhere, a substantial body of dissenters exists. It is unlikely in such cases that opinion will ever be defeated by fact.

As a stimulus to intuition in applied research. For analysts concerned about delivering a timely and cost effective solution, the correct criterion is efficiency: the most good ideas for the resources consumed. Instead of trying to assign dollar values to ideas, which is probably impossible, consider comparing the bad model to another source of ideas. For concreteness, consider a slightly facetious alternative to combat models. A research team could mock up several sets of notional briefing charts, with blanks on the charts in place of numbers that the model might produce. They could then hire my 15-year-old nephew to write numbers in the blanks. For each set of charts, the research team would try to devise a plausible explanation for the ensemble of numbers, and if none was found, the set would be discarded. If a plausible explanation were devised, the researchers would be in the same logical position as if they had gotten the numbers from a combat model: My nephew's charts had suggested, the truth must be found elsewhere. The efficiency comparison is straightforward. Detailed combat models come at a high cost in person-hours, hardware, and software. My nephew will probably produce more unusable charts than the model, but he is far cheaper per set of charts.

One might object that, unlike my nephew's numbers, model output is motivated by the substantive content of the model's assumption. Perhaps; it depends. For combat models, validations usually consist of checking whether model output is reasonable, and if it is not, the model is changed. The best combat modelers make gross comparisons between model output and, e.g., historical attrition rates, but if the output is not reasonable—if no plausible explanation can be constructed—the model is changed. The exercise with my nephew is much like the actual process, and in this sense the substantive content of model assumptions is oversold.

Some have argued that a model is a record of past learning. Again, perhaps. For combat models, this is true in some particulars, but as an aggregate the model is a record of created outputs that were susceptible to a plausible explanation, *which does not imply that anything has been learned about the world.* Consumers take note.

But now we come to the sales benefits of a model used as a stimulus to intuition. A model sells like my nephew never will. Why else use an expensive model to generate output when the output cannot legitimately be more than an illustration, and a much cheaper illustration can be had just by inventing numbers that illustrate the same thing? This is the bite of the stricture that a bad model can only suggest, not reveal: If information content were the criterion, we would rarely lose anything by banning model output from briefing charts. But we dare not, because models sell. In particular, for some clients superficial realism is immensely important in building up the probability of a sale. This is why the program manager judged that it does little harm to sell a number as a model output if it is a good number: better to sell a good number with some meaningless flash than not to sell it at all. But clients pay—a lot—for this flash, and if it is necessary only because of the clients' own cultural predispositions, they might be better off if they shed those predispositions.

5. **As an aid in selling an idea of which the model is but an illustration.** All that is required here is to get the idea across. Superficial realism may be essential, for example, if the potential buyer is going to get into the cockpit and pull levers on the model. In such cases, the expense of realism and complexity may be unavoidable. But it might be enough simply to have briefing charts with the right labels for inputs and outputs and a story, but not necessarily specific computer code, that connects them. In this case, selling can be cheap. It need not involve deception, as long as the clients know what they are being sold, the sales materials are only used for illustration, and the analysts believe they can deliver.

6. **As a training aid to induce particular behavior.** Cost-benefit considerations here are much the same as under "as a stimulus to intuition in training:" If possible, measure whether the behavior has, in fact, been induced, and measure the cost of inducing it.

3. CONCLUSION

This paper has two messages, one pragmatic and the other theoretical. The pragmatic message is for clients: Analysts are often vague about the actual purpose served by a model, and if they are forced to be clear, the model's contribution often turns out to be less than is apparent and overpriced for what it delivers.

The theoretical message is aimed at analysts, and is the framework and approach of the paper. The theory presented here is obviously incomplete. When it is more complete, we may hope for a solid theory of bad models. This is not to suggest a formal theory of costs and benefits, with equations and numbers—that would be a step toward infinite regress and a waste of time. Ideas embodied in words should suffice.

This paper may seem uncharitably hard on people who toil long hours to shed light on difficult problems. As one colleague put it, "sometimes a very rough answer to a difficult problem is better than no answer at all—the trick comes in knowing when this is so!" True enough, but this leaves the question of the cost effectiveness and logical candor of different ways to reach the rough answer, and the question of when that answer is good enough. One might hope that the approach taken here will help analysts and consumers answer these questions, and that we will hear less of the evasion that modeling is an art or a craft.

ACKNOWLEDGMENT

In writing this paper I have had the benefit of comments from several RAND colleagues: Steve Bankes, Bart Bennett, Carl Builder, Jim Dewar, Fred Frostic, Glenn Gotz, Emmett Keeler, Jacob Klerman, John Nachbar, Ray Pyles, Elizabeth Rolph, John Rolph, Norm Shapiro, Jack Stockfisch, and Warren Walker; and from David Freedman of the University of California at Berkeley. Peter Stan contributed some useful formulations. Paul Davis and Bob Levine made particularly helpful comments. Hugh Miser contributed to a remake of some of the validation material, and he and Gene Woolsey made valuable editorial suggestions. These generous people do not necessarily agree with anything in this paper. The views expressed in this paper are the author's own and are not necessarily shared by RAND or its research sponsors.

REFERENCES

COHEN, I. K., J. B. ABELL AND T. F. LIPPIATT. 1991. Coupling Logistics to Operations to Meet Uncertainty and the Threat (CLOUT): An Overview. RAND R-3979-AF, Santa Monica, Calif.

DEWAR, J., J. GILLOGLY AND M. JUNCOSA. 1991. Non-monotonicity, Chaos, and Combat Models. RAND R-3995-RC, Santa Monica, Calif.

HODGES, J. S. 1989. Building the Brooklyn Bridge, Then Buying It. Unpublished manuscript.

MISER, H. J. 1988. Underlying Concepts for Systems and Policy Analysis. In *Handbook of Systems Analysis: Craft Issues and Procedural Choices*. North-Holland, New York, 467–525.

MISER, H. J., AND E. S. QUADE. 1988. Validation. In *Handbook of Systems Analysis: Craft Issues and Procedural Choices*. North-Holland, New York, 527–565.

THOMAS, C. J. 1989. Verification Revisited—1983. In *Military Modeling*, 2nd ed., W. P. Hughes, Jr. Military Operations Research Society, Alexandria, Va., 255–271.

RAND/N-3381-RC