CHOOSING HOW TO CHOOSE: COMPARING AMALGAMATION METHODS FOR ENVIRONMENTAL IMPACT ASSESSMENT

BENJAMIN F. HOBBS

Amalgamation, in which disparate impacts are combined so that alternatives can be ranked, has become an important part of many impact assessments. Such methods can help make decisions more rational by systematically combining great amounts of information into more digestible forms. They can also facilitate public participation and ease documentation of decisions. The intent of this article is to give an overview of amalgamation methods and to propose four criteria for choosing among them: the purpose to be served, ease of use, validity, and results compared to other methods. Because experiments have repeatedly shown that the method chosen can significantly affect what decision is made, EIA practitioners should place more emphasis on the last two criteria than they have in the past. Finally, recent results in psychology and management science are discussed for practitioners facing the question "how do we choose how to choose?"

The author is Assistant Professor of Systems Engineering at Case Western Reserve University, Cleveland, Ohio. He has been working in the area of multiple criteria decision making since 1975 at Oak Ridge and Brookhaven National Laboratories, the Maryland Power Plant Siting Program, and the College of Environmental Science and Forestry, State University of New York, Syracuse, where he earned an M.S. in 1978. He obtained his Ph.D. in Civil and Environmental Engineering from Cornell University in 1983. He is presently conducting a risk-benefit analysis of oil and gas brine disposal and developing methods for assessing the reliability of urban water systems.

Environmental impact assessment (EIA) includes a bewildering variety of seemingly incomparable impacts. For example, an impact statement for a proposed power plant might describe its effects upon river temperature, endangered species, worker immigration rates, and the scenic qualities of the surrounding area.

In combining such diverse considerations into an evaluation of worth, unaided human judgment is notoriously unreliable (Shepard 1964; Hammond et al. 1975). The mind may flit inconsistently among different aspects of the problem or become fixed on one or a few attributes and disregard the remaining (but still important) ones. Furthermore, those who make such unaided judgments are generally unable to recall how they considered the combined impacts and what importance they assigned to each (Nisbett and Wilson 1977). Amalgamation methods, also known as multicriteria decision-making techniques, can be useful in these circumstances. They can help document evaluations and make them more reliable and consistent with the values of decision makers.

The purpose of amalgamation techniques is to commensurate the incommensurable. They combine impacts so that alternatives can be ranked according to their "desirability" or "total impact." A single numerical index of worth is often the result. As examples, concentrations of different air pollutants can be combined into a measure of air quality, or, as in many power plant impact statements, all the environmental, social, and economic attributes of an alternative might be amalgamated into an index of "suitability." This process reduces information overload, and can make it easier for decision makers to focus systematically upon the most important tradeoffs.

Difficult value judgments are *inevitably* required when amalgamating "facts" (attributes of alternatives that, in principle, can be measured objectively) into an index or a ranking of alternatives. For example, the combination of different types of water pollutants, each of which has a different kind and degree of effect, into an index of water quality is not simply a factual decision, since "water quality" is not objectively measurable. Such an amalgamation requires the making of subjective value judgments about health, aesthetic, ecological, recreational, and cost factors. Thus, amalgamation is inherently subjective and value laden. It must be noted that amalgamation does not pollute an otherwise "objective" and "scientific" impact assessment with subjectivity; rather, it makes the inevitable value judgments explicit.[1]

Explicit amalgamation of impacts has become an important part of many impact assessments, and a wide variety of amalgamation techniques have been proposed and applied in EIA.[2] These methods differ in terms of the mechanics of attribute combination, the assumptions they make about decision makers' values, and their purposes and ease of use. There is mounting evidence that the alternative chosen often depends on the amalgamation method used. Those who

[1]Several writers in the *EIA Review* have stressed the importance of recognizing the value judgments that pervade EIA, of which amalgamation is merely one of many (Bacow 1980; Mauheim 1981; Susskind and Dunlap 1981).

[2]Approximately half of the EIA methods described by Henderson (1982) perform amalgamation using value judgments. Pierce and Rowe (1979), for the case of nuclear power plant impact statements, and Canter (1979), for municipal wastewater program impact statements, cite numerous examples where amalgamation has been used in EIA. EIA applications of a variety of amalgamation methods are summarized by Elliot (1981), McAllister (1980), Nichols and Hyman (1982), and Hobbs *et al.* (1984). Nonetheless, Bisset (1980) feels that amalgamation will be, and should be, deemphasized in the future in favor of a more political style of decision making.

undertake EIAs should therefore carefully consider which amalgamation technique is most appropriate for their situation. The purpose of this paper is to discuss some considerations involved in picking an amalgamation method and to review recent results in psychology, management science, and other fields for EIA practitioners facing the question "how do we choose how to choose?"

To pick a method, one needs to know what different techniques there are, by what criteria they might be judged, and how the methods perform relative to those criteria. After presenting a brief overview of available methods, four criteria are proposed: the purpose served, ease of use, validity, and the effect of the choice of method. Some relevant theoretical and empirical results from the literature of multicriteria decision making are then summarized. These results clarify the relative advantages of some of the methods and suggest when choice of method might matter.[3]

APPROACHES TO AMALGAMATION IN EIA

Only a brief review of available amalgamation methods for EIA can be presented here. There are a number of descriptions of the methods and their assumptions and guides to their implementation in EIA (e.g., Elliot 1981; Brown et al. 1980; Lord et al. 1979; Hobbs 1979; McAllister 1980; Bakus et al. 1982; Henderson 1982; O'Banion 1980).

Amalgamation serves two main purposes in EIA; prescription and description. Prescription is the role most often played; the object in that case is to indicate which alternative(s) should be preferred, given the options available, their impacts, the values of the decision makers, and presuppositions as to what constitutes rational decision making. For an amalgamation method to yield a decision fully consistent with a decision maker's preferences, it is necessary, among other things, that (1) the attributes completely describe the important impacts and characteristics of the alternatives, while avoiding redundancy, and that their values or probability distributions for each alternative be known, (2) all relevant alternatives be included, and (3) the decision makers have a stable set of preferences and be able to voice them (see Keeney and Raiffa 1976). The first two frequently cause controversy in EIA.[4] The latter requirement is of particular concern here since it is likely to be unsatisfied in impact assessment and its

[3]Amalgamation methods are discussed in this paper as though a single decision maker makes all value judgments. This is of course a gross simplification, as EIA involves not only professionals with diverse backgrounds, but also the public. There exist many methods, such as Delphi and the Nominal Group Technique, by which value judgments such as weights can be obtained from groups; Bakus et al. (1982) discuss their use in EIA. Space does not permit discussion of these methods and their pros and cons. Nonetheless, the criteria proposed here for evaluating amalgamation methods are also relevant to multidecision-maker techniques.

[4]For example, witness the current controversy over the procedure used by the U.S. Environmental Protection Agency to rank hazardous waste dumps for clean up (Wu and Hilger 1984). Critics quarrel with the rankings, asserting that the list of attributes used is incomplete and that important data is absent.

violation means that the particular amalgamation technique chosen can strongly bias the values voiced.

Amalgamation is also sometimes used to describe and predict choices in EIA. For example, by modeling how various interest groups perceive alternatives, intergroup conflict can be better understood and perhaps resolved (see Dennis et al. 1983; Brown 1984). Descriptive models are often developed by using multiple regression or linear programming to relate attributes to a decision maker's overall evaluations of the worth of alternatives. Note that prescribed and predicted choices may differ substantially, even for the same person. The reason is that prescriptive methods try to *improve* judgment by imposing certain rules that are assumed to characterize rational decision making based upon the decision maker's values; descriptive methods attempt to *imitate* subjective judgment, warts and all.

Alternative amalgamation methods are summarized below in terms of how they accomplish each of the following tasks: (1) attribute scaling, where a single impact (e.g., "ppm NO_x" or "families displaced") is converted into a measure of value and, in some applications, the decision maker's attitudes towards risks are captured; (2) attribute weighting, where each attribute's importance is assessed; and (3) amalgamation of the weighted and scaled attributes via a decision rule, yielding an indicator of overall value or a ranking of alternatives. Not all amalgamation methods include every step, while several include additional ones. As an example of the latter, a separate amalgamation might be performed for each of several interest groups, followed by a fourth step which combines those perspectives into a single evaluation.

Another step that all applications should include is a verification of the assumptions the chosen method makes about the structure of the decision maker's values. For example, the New York State Department of Environmental Conservation (1980) used a weighted sum of scaled attributes to evaluate alternative sites for hazardous waste treatment facilities. Such a method presupposes a property called preference independence: preferences among combinations of levels of any pair of attributes do not depend upon levels of other attributes. However, this property does not hold for many of the attributes used in that siting study. For example the tradeoffs someone would be willing to make between "population density in the vicinity of the site" and "impact on endangered species" would probably depend on the "risk of fires and explosions at the site." In this case, defining a new attribute "population at risk times risk of fires and explosions" might rid the attribute set of this interdependency. However, when violations of independence assumptions are not so easy to deal with, the question becomes: is a more complex amalgamation method appropriate? The answer depends on the relative ease of the alternative methods and whether a practical difference would be made by picking a different method. Later sections of this paper address this question.

Scaling Methods

Scaling can accomplish two goals: the translation of an attribute into a measure of value and the measurement of the decision makers' attitudes towards risk. Value scaling methods do only the former; utility scaling methods accomplish both. Utility methods are preferred if: risks are measured in the form of a probability distribution for each impact for each site, decision makers are not risk neutral (i.e., value and utility methods yield different results), and decision makers can meaningfully and consistently respond to the questions posed by utility methods. Most EIA applications adopt value scaling.

Because addition or multiplication of arbitrary ordinal scales is not mathematically permissible, value scaling methods which yield functions that are interval or ratio scaled with respect to value should be used.[5] That is, differences in numbers should be meaningful. An example of a value scaling question is (Dyer and Sarin 1979): "what level B of the attribute would make a change from level A to B just as desirable as a change from B to level C?" A less rigorous approach would be to make an impressionistic sketch of a function relating the attribute to value. Another deterministic method, conjoint measurement, scales two attributes simultaneously by asking questions about the tradeoffs the decision maker is willing to make between them (Keeney and Raiffa 1976). Yet another approach is holistic scaling, in which the desirabilities of different levels of an attribute are inferred using, e.g., multiple regression from unaided overall evaluations of alternatives.

Utility scaling methods ask decision makers to state the probability p (or level B) at which they are indifferent between one alternative having a chance p of obtaining level A and chance $(1\text{-}p)$ of level C, and a second alternative having level B for certain. The utility of B would be $pu(A) + (1-p)u(C)$, where $u(\cdot)$ is the single attribute utility function. Such a utility function might be nonlinear even if the associated value function is linear because the DM may either dislike or prefer gambles. Examples of the use of risky utility functions in an environmental quality indices include Collins and Glysson (1980) and Keeney and Robilliard (1977).

Weighting Methods

Weights are the means by which many amalgamation methods determine how much of one attribute the decision maker is willing to give up for another. For example, in the weighting summation decision rule, attributes are multiplied by weights and summed, yielding a single index. Under that rule, if a DM is just willing to give up one unit of scaled attribute i in order to obtain an improvement

[5] Certain types of ordinal scales can be added or multiplied, such as those yielded by conjoint measurement (Keeney and Raiffa 1976; Dyer and Sarin 1979). However, this is not true with arbitrary ordinal scales.

of two units in scaled attribute j, the weight w_i of attribute i must be twice that of attribute j. If $w_i/w_j \neq 2$, then the rule will incorrectly show that the DM either is willing to give up more than one unit of i for two of j or is unwilling to give up even that one unit, either case being a distortion of the decision maker's preferences. The result can be a decision that is not really preferred based on the tradeoffs the decision maker is willing to make.

Several approaches to weighting exist. Some ask decision makers to directly assess each attribute's "importance," while others infer weights from tradeoffs the decision maker is willing to make. A third approach solves for the weights that best imitate unaided evaluations of alternatives.

Direct assessment methods, or "magic number" methods, are the most frequently used approach in EIA (e.g., Brown 1984). In one version, called "rating," the decision maker directly chooses numbers on a scale representing the relative "importance" of each attribute. Variants of the magic numbers type of method include: (1) ranking of attributes, designating the lowest ranked attribute a weight of 1, the next lowest a 2, and so forth; (2) ratio questioning, in which one is asked the ratio of the "importance" of two attributes; and (3) allocation of 100 points among attributes. The assumption of magic number methods is that the ratio of the "importance" of two scaled attributes will, in the decision maker's mind, correspond to the reciprocal of the rate at which the decision maker is willing to trade them off. Although some psychologists argue that this correspondence exists (John and Edwards 1978), others point out that "importance" could mean many different things, and does not necessarily represent willingness to trade off (Hobbs 1980; Schoemaker 1981).

The second approach to weighting asks decision makers what tradeoffs they would be willing to make among attributes. If the decision maker is asked how much of one attribute he or she is willing to give up to obtain a given amount of another, the method is termed the indifference tradeoff technique. The gamble method, in contrast, asks about tradeoffs by requesting the decision makers to specify at what probability p they are indifferent between: (1) an alternative with chance p of a set of attributes being at their best level and chance $(1-p)$ of them being at their worst; and (2) an alternative with one of those attributes at its best level (whose weight will be proportional to p) and the others at their worst. Collins and Glysson (1980) applied both these methods in an EIA.

The third approach, the holistic method, chooses weights to maximize the predictive value of the resulting amalgamation. An example of an EIA using a regression holistic method is that of Hyman and Stiftel (1980).

Decision Rules

Decision rules combine scaled attributes and weights into a single index or ranking of alternatives. Like scaling methods, decision rules can be constructed so that they represent the relative value of alternatives, or they can also include

the decision maker's attitudes towards risk. In the latter case, they are called multiattribute utility functions. The set of decision rules can also be sliced another way. Noncompensatory rules do not allow improved performance in one attribute to make up for bad performance in another. Compensatory rules do permit such compensation. Compensatory rules can be divided into linear and nonlinear (or configural) rules. In the linear case, it is assumed that the rate at which the decision maker is willing to trade off any scaled attribute for any other depends neither on the levels of those attributes nor on the levels of the remaining attributes. Other assumptions concerning attribute independence are made by configural rules.

Two simple decision rules that find frequent use in EIA are exclusionary screening and weighting summation (Elliot 1981). For example, Pierce and Rowe (1979) cite eight nuclear power plant impact statements that used the former method to help choose suitable plant sites and 13 that applied the latter. Exclusionary screening, a noncompensatory rule, merely drops any alternative that is not satisfactory with respect to each attribute. "Satisfactory" can be defined using inviolable legal or engineering criteria. Alternatively, it can be specified more arbitrarily for discretionary attributes (e.g., prime farmland consumed or distance to water) when the purpose is to quickly narrow the list of alternatives. Weighting summation, in contrast, is a linear compensatory rule.

One of many nonlinear compensatory rules is compromise programming. Its premise is that decision makers prefer to be as close as possible to an ideal point. Distance is defined as the sum of weighted deviations of attribute levels from goals, each deviation raised to some predefined power between one and infinity.

Another nonlinear compensatory rule is the multiplicative rule:

$$1 + kv(\mathbf{x}) = \prod_i (1 + kw_i v_i(x_i)),$$

Where the $v_i(x_i) \in [0,1]$ are the scaled attributes, $w_i \in (0,1)$ are the weights, and $k \in (-1, \infty)$ is a constant which ensures that the overall value $v(\mathbf{x})$ of an alternative $\mathbf{x}$ falls in the range $[0,1]$. If $k = 0$, then the multiplicative rule reduces to weighting summation. Keeney and Raiffa (1976) developed and justified this rule for the risky decision/utility theory case; Dyer and Sarin (1979) did the same for the riskless decision/value theory case. The rule's independence assumptions are slightly weaker than those for weighting summation. These approaches stand at the apex of a body of theory whose completeness and elegance are without parallel in multicriteria decision making. For this reason, many practitioners labor through the assumption verification, scaling, and weighting procedures necessary for their valid use—even in situations where less difficult methods would yield the same evaluations. An EIA application of the multiplicative rule is given by Collins and Glysson (1980).

The power law is another nonlinear compensatory rule, in which weights are applied as exponents to scaled attributes and the attributes are multiplied together.

The assumption is that decision makers perceive "importance" logarithmically; that is, for example, if a decision maker says that two attributes should receive weights of one and two, respectively, then their actual relative importance is e^1 and e^2. Yet another rule, concordance analysis, is simultaneously compensatory and noncompensatory. In one version, ELECTRE, an alternative is said to outrank another if it is better in a sufficiently large number of attributes and not too much inferior in any of the other attributes. The philosophy of this approach can be characterized as being one of "majority rule," subject to protection of the "rights of minority viewpoints." Gum et al. (1976) used the power law to amalgamate environmental attributes. The use of concordance analysis in EIA was suggested by Bakus et al. (1982) and an application is found in Gershon and Duckstein (1983).

CRITERIA FOR SELECTING AN AMALGAMATION METHOD

What criteria should be considered when choosing an amalgamation technique for use in EIA? Four are proposed here: the purpose served, ease of use, validity, and results compared to other methods. In general, no single method will be superior in all criteria for a particular use—thus, picking an amalgamation method is a multicriteria problem in itself! Techniques adequate in some aspects will be deficient in others. As an example, a magic numbers weighting method may be simple to apply, but the resulting weights might be invalid as they may bear no relationship to tradeoffs decision makers are willing to make. The significance of this possible invalidity depends on the question implied in the last criterion: would more valid methods lead to the same decision?

A problem for potential users of amalgamation is that the number of methods is huge, and the associated literature is voluminous and growing rapidly. Although several texts survey the field (e.g., Chankong and Haimes 1983; Cohon 1978; Goicoechea et al. 1982), none explicitly compare methods relative to the above criteria. Experts disagree about which criteria are most important and how the methods actually perform against them. Further, as Warner and Preston (1974) point out, method characteristics that are virtues in some EIA problems may be vices in others.[6] For these reasons, one cannot blame EIA practitioners if they throw up their hands and decide to use whatever method happens to be the most convenient.

In the remainder of this section, the four criteria are discussed in some detail. The following section summarizes several recent research results that clarify how different techniques meet the criteria.

[6] Janssen and Nijkamp (1985) offer a typology of environmental management problems, a listing of the roles that amalgamation should play in each problem, and a summary of the appropriateness of available methods for each problem. In doing so, they emphasize two of the four criteria of this paper, "purpose" and "ease of use."

Criteria I: The Purpose Served

Amalgamation plays several roles. The major one is to condense large amounts of information into a form that makes it easier to make a rational decision. This form could consist of a small set of indices, each of which summarizes impacts in one category (e.g., a water quality index could include turbidity, inorganic and organic pollutants, dissolved oxygen, and pathogens). The tradeoffs that exist among altenatives may become more apparent when so distilled; how much amalgamation should be made depends upon the point at which further simplification hides rather than explicates tradeoffs. Weighting summation, for example, can be appropriate for such a use, but not exclusionary screening or concordance analysis. The latter methods can only rank alternatives or group them into categories (accept/reject). They could be suitable, however, if the purpose is instead to make a single overall evaluation of the alternatives. Yet, even in that case, a method such as weighting summation might be preferred because it leads to a quantitative indication of each alternative's worth.

Whether a quick screening of undesirable alternatives or a carefully considered and documented choice is desired is also important. In the former case, a simple technique may be satisfactory; in the latter, a method which is more likely to be valid may be better suited. If the most important "product" of an analysis is the process itself, then a transparent, easily applied method is probably best. This may be the case, for example, when public participation in an EIA is desired as an end in itself.

If one wishes to better understand the values held by interest groups and to predict their evaluations, then a method that imitates unaided judgment might be preferred to one that uses a "divide and conquer" approach to improve it. An example is multiple regression (Hammond et al. 1975).

Criterion II: Ease of Use

The resources available to a typical EIA restrict the sophistication with which an amalgamation can be carried out. However, it does not follow that it is desirable to minimize the effort needed to implement a method, because to do so might endanger understanding and acceptance of the method by decision makers and the public.

The ease of applying a given method depends on the time needed to acquaint users with the technique, how comfortable they are with its concepts and questions, and the time and computer facilities it requires. These depend, in turn, upon (1) what knowledge the analysts and decision makers possess and (2) the complexity of the problem, including numbers of alternatives and attributes, whether attributes are measured quantitatively, the degree of aggregation desired, the number of decision makers, and how much their values agree.

Criterion III: Validity

Validity refers to how well a technique measures the concept it purports to measure. Thus, for example, a magic numbers weighting method can be invalid because decision makers may not be thinking of the tradeoffs they are willing to make when rating each attribute's "importance." In that case, the method is not measuring the correct concept. The result can be a decision that is inconsistent with the decision maker's values. Applications of amalgamation in EIA have often been criticized as being invalid (Hobbs 1979; Adelman and Mumpower 1978).

Measuring an incorrect concept during scaling or weighting is but one possible source of invalidity. Another important one is when a decision rule is inappropriate for the decision maker's value structure; e.g., if a compensatory rule is used where a noncompensatory one is called for. A similar problem occurs if the decision maker's unaided choices violate the fundamental axioms of the adopted amalgamation method, such as transitivity. When amalgamation is used for prescription, this type of invalidity may actually be desired, because one wishes to strip human judgment of its irrationality. However, violations of axioms usually imply that what scales and weights are elicited from a decision maker will depend on supposedly arbitrary aspects of question phrasing (e.g., using 50–50 probabilities in gamble scaling versus using 75–25 probabilities). This lack of uniqueness means that the analyst or decision maker must decide which sets of scales and weights are most appropriate—when there are no clear criteria for making such a decision.

Hyman (1981) describes other sources of invalidity. One type of bias, hypothecality bias, occurs when a person's actual choices fail to coincide with his or her answers to a method's hypothetical questions. Another type, strategic bias, occurs when people choose not to express their true preferences. Instrumental bias is a problem when seemingly unimportant aspects of a method, such as question phrasing or ordering, affect the results. Fischhoff et al. (1979) argue that this is likely to be a severe problem when decision makers are unsure of what they want. This, in turn, is probable when problems are nonroutine, complex, and involve strongly held but conflicting values. EIA problems fall into this category.

Invalidities are likely to minimized if: (1) scaling and weighting questions are phrased so that they directly measure the correct concept; (2) questions are framed to check whether the decision rule is appropriate; (3) unfamiliar concepts and hypothetical choices are avoided; (4) the method is thoroughly understood by users; (5) consistency checks are made; (6) decision makers are familiar enough with the problem and with their own values so that they know what they want; (7) incentives for honest responses are present; and (8) more than one set of scaling and weighting techniques are used and their results are similar.[7] Time

[7] For example, Dennis et al. (1983) had participants in an evaluation of air quality management alternatives use more than one weighting method in order to avoid the bias that might result from using a single method.

and resource constraints normally prevent fulfillment of every one of these characteristics in an EIA. However, to the extent that they can be satisfied, one can be confident that the resulting amalgamation truly reflects values held by decision makers.

Criterion IV: Results Compared to Other Methods

In general, methods that are more likely to be valid also require more effort. That extra effort can only be justified if the more valid method has some probability of leading to a different decision. If it doesn't, one shouldn't bother. Many psychologists have, for example, argued that it doesn't matter whether differential weights are used when attributes are positively correlated; any set of weights, including equal ones, will yield roughly the same rank ordering of alternatives. However, a growing body of evidence, summarized below, indicates otherwise: which amalgamation method is chosen can make an important difference. If this is so in a particular EIA application, then the EIA study team should carefully consider which method is most likely to provide valid representations of the values of the decision makers and the public.

The results of different methods can diverge for several reasons. Perhaps the most important one is that decision makers are not sure of their values; when this is true, instrumental biases can cause divergence of results (Fischhoff et al. 1979). Other reasons include incorrect choice of decision rule; weighting or scaling methods that measure the wrong concept; and user fatigue, boredom, and, particularly, misunderstanding. Perceptual biases can also be important. For example, magic numbers weights tend to be much more uniform among themselves than weights based upon tradeoff questions or regression analysis of unaided judgments (Nisbett and Wilson 1977; Hobbs and Rowe 1979; Schoemaker and Waid 1982; John et al. 1980). One possible explanation is logarithmic perception, explained above. Another is that people tend to avoid extreme numbers when rating, preferring values in the middle of scales. A third possible explanation for the relative flatness of magic numbers weights compared to regression weights is that holistic judgments exaggerate the importance of the two or three most important attributes. This occurs because the complexity of a holistic judgment causes decision makers to simplify the problem by considering only a handful of attributes and ignoring the remaining (though still important) ones (Shepard 1964). Such perceptual biases will cause the results of different methods to diverge.

A necessary condition for divergence is that there be tradeoffs among the alternatives. If the tradeoffs are many, then small changes in a decision rule or its parameters could cause a large shift in the decision.[8] EIA problems are often characterized by many and important tradeoffs.

[8] McClelland (1979) and Rowe and Pierce (1982a) have systematically investigated this "decision complexity" and how it affects the influence that weights have on a decision.

SOME RECENT RESULTS

Within the last half decade, researchers in multicriteria decision making have obtained theoretical and empirical results that clarify how several amalgamation techniques perform on the four criteria outlined above. One important theoretical result is the completion of a firm theoretical base for the type of amalgamation technique underlying several EIA methodologies. A number of experiments have also been made which indicate when choice of method is likely to make a difference. These and other results are summarized below.

Scaling Methods

Recent work in scaling has focused on (1) comparing the results of deterministic/value and risk/utility scaling methods and (2) the validity of the expected utility model which underlies utility scaling.

A number of experiments have discovered persistent differences between single attribute value functions chosen by deterministic scaling techniques and utility functions chosen by the gamble method (Krzysztofowicz 1983). Sarin et al. (1980), for example, found this to be true for attributes measuring power plant impacts upon air quality, fish, site biology, socioeconomics, and environmentally sensitive areas. They and Krzysztofowicz (1983) explained such differences by saying that utility functions include both strength of preference for levels of an attribute and risk attitudes, while value functions include only the former. The different scaling methods used in that experiment nevertheless yielded the same ranking of power plant sites (Sarin 1980). Thus, choice of scaling technique seems to have less effect on the overall results than choice of weights, which in that study did shuffle site ranks significantly.

Psychologists have conducted many experiments that show that people make choices under risk in a manner inconsistent with the expected utility model (Schoemaker 1982; Fischhoff et al. 1980). This has two implications for the EIA practitioner. The first is that for this reason alone an amalgamation based on these axioms can be expected to differ from holistic evaluations of alternatives. The second implication is that the gamble method will be unreliable. This is because the resulting utility function will depend strongly on aspects of gamble questions which, according to the axioms, should be irrelevant. For example, utility functions based upon gamble questions using 50–50 probabilities will differ from those resulting from 70–30 probabilities. In addition, "noise" will be introduced because many decision makers will be unfamiliar with the laws of probability and uncomfortable with gamble questions. For these reasons, it can be argued that value scaling methods are more appropriate for EIA, unless (1) risks in each alternative's attribute levels are explicitly specified using probability distributions and (2) utility functions are likely to differ significantly from value functions (O'Banion 1980; Hobbs 1979).

Weighting Methods

A number of researchers have asked: to what extent do different weighting methods measure the correct concept (tradeoffs that decision makers are willing to make) and yield the same results? Hobbs (1980) and Schoemaker (1981) have presented theoretical reasons why magic numbers methods are likely to measure the wrong concept, and McClelland (1979) and Newman (1977) have convincingly demonstrated that different weights are likely to yield different decisions. But evidence from studies where decision makers have applied several weighting methods is needed to determine which methods are likely to differ under what circumstances.

At least 10 studies have contrasted the results of different magic numbers methods, such as point allocation, rating, and ranking. They unanimously conclude that the differences in weights and the resulting evaluations of alternatives are minor compared to, for example, differences between evaluations by different people (e.g., Hobbs and Rowe 1979; Hobbs 1985).

Several studies have compared magic numbers techniques with indifference tradeoff and regression methods. Some found a high convergence among the resulting evaluations of alternatives (e.g., John et al. 1980; Schoemaker and Waid 1982) and conclude that choice of weighting method matters little, while others find important disagreements in weights and, frequently, the ranks of alternatives (e.g., Nisbett and Wilson 1977; Saaty et al. 1983; Hobbs 1985). For example, important differences in decisions were found in a power plant siting study which involved six cost-related attributes and 12 environmental attributes (Hobbs 1980). Five experts in siting applied both rating and the indifference tradeoff methods. The weights were used to screen out 92 percent of a study region, leaving areas that would be suitable for later, more detailed investigation. On average, just half of the areas chosen by a person by one method were also chosen by the second method. This is less than the mean overlap between areas chosen by different people using rating (62 percent). The conclusion was that choice of method was at least as important as who chose the weights.

Disagreements in the conclusions of studies that compared different weighting methods appear to stem at least in part from the characteristics of the decision problem. When (1) decision makers are very familiar with the problem, (2) there are only three or four attributes, and (3) where conflicts between strongly held values (such as the desire to preserve the environment versus the desire to keep utility rates down) are absent, then convergence among methods is found and indeed expected. However, unique decisions, many attributes, or conflicts among important values can lead to divergence. Divergence occurs because (1) people have not decided exactly what tradeoffs they are willing to make and magic numbers methods fail to force them to consider tradeoffs, and (2) people cannot consider the full complexity of the problem when making the holistic choices required by regression weighting. Decisions in EIA are often one of a kind, possess many attributes, and involve strong value conflicts. Hence, it should be assumed that the choice of a regression, magic numbers, or indifference tradeoff technique will affect which alternatives are chosen, unless proof to the contrary is provided by application of more than one method.

Decision Rules

Two recent research topics concerning decision rules are particularly relevant to the practice of EIA. The first concerns the relationship of multiattribute utility and value functions. The second is the question: does choice of decision rule make a difference that matters?

Developments in utility theory in the early 1970s provided a rigorous theoretical base and practical set of procedures for constructing multiattribute utility functions for decision making under risk (Keeney and Raiffa 1976). In theory, such functions should also yield the same ordering of riskless alternatives as valid value functions (which exclude attitudes towards risk). The advantage that utility theory used to hold in the deterministic case was that multiattribute utility functions could be constructed from single attribute utility functions derived one at a time using the gamble method. At that time, valid multiattribute value functions could only be built from single attribute value functions that were estimated by conjoint measurement, defined above, where two ordinal single attribute value functions are estimated simultaneously using different tradeoff questions (Keeney and Raiffa 1976).

At that time, no theoretical basis existed for independent creation of single attribute value functions, such as those in the Environmental Evaluation System (Dee et al. 1973), and their assembly into a multiattribute value function. Dyer and Sarin (1979) established such a basis. Their requirements for theoretical validity of a weighting summation or multiplicative "measurable" value function are (1) preference independence among attributes and (2) that each attribute's value function be interval-scaled and independent of the levels of the other attributes. This theory is important to EIA because it provides a defensible theoretical basis for building EIA methodologies from independently determined attribute value functions.

Several experiments have been conducted to determine when choice of decision rule makes a difference in alternative ranks. Some have compared the multiplicative rule under various values of k (which determines the degree of interaction) with weighting summation. When the same relative sets of weights are used, most experiments have found that if k is kept within reasonable bounds (roughly -0.5 to 2.0), then choice of k, including $k = 0$ (weighting summation) does not significantly affect the decision (Hobbs 1979; Collins and Glysson 1980).

Comparisons have also been made of other decision rules. A number of studies have demonstrated the "robustness" of weighting summation by showing that it predicts subjective decisions as well as other, more complex decision rules

(Fischhoff et al. 1980; Hobbs 1985). Nevertheless, there are occasions when complex rules are more appropriate and lead to different decisions. A number of studies have shown how choice of rule can matter. Weighting summation, the power law, exclusionary screening, compromise programming, and ELECTRE resulted in different decisions in most, but not all hypothetical power plant siting studies (Hobbs 1979; Hobbs and Rowe 1979; Hobbs 1985; Rowe and Pierce 1982b). Gershon and Duckstein (1983) applied compromise programming, multiplicative utility functions, and ELECTRE to a river basin planning problem, and again choice of method significantly affected the ranks of alternatives. Indeed, the correlation between the two sets of alternative ranks chosen by the utility function and ELECTRE did not significantly differ from zero. However, the conclusions of these studies are necessarily of limited scope, as most of them did not systematically examine the separate effects of weighting method, decision rule, and complexity of decision. All that can be concluded is that choice of decision rule can make a difference that matters, and that EIA practitioners should carefully consider which rule is most appropriate for their particular situation.

CONCLUSION

Amalgamation methods can serve useful roles in EIA. They can simplify decisions and help make them more rational by combining large amounts of information on alternatives into a more digestible form. They also ease decision documentation and can facilitate public participation. However, there are dangers. Because (1) decision makers often don't know what they want, (2) different people hold different values, and (3) choice of method can affect the decision, the results of amalgamation techniques should be taken with a pinch, perhaps even a shaker full of salt. The precision of their numerical evaluations should be mistaken neither for accuracy nor for concensus. The writer agrees with Stiftel and Hyman (1980) and Hollick (1981) who argue against letting "super-indices" make what are, and should be, political decisions in EIA. More appropriate roles for amalgamation in EIA include (1) the summary of information *within* categories of impacts (e.g., air quality or socioeconomics) and (2) the representation of perspectives of different interest groups, each of which is relatively homogeneous. Amalgamation techniques should clarify tradeoffs and value conflicts, not hide them (Elliot 1981).

If an amalgamation method is to be used to produce a final ranking of alternatives, it should be applied with utmost care and with the realization that choice of method might affect the decision. The prudent course for those who use amalgamation for any purpose is to (1) check assumptions, (2) use more than one method, and (3) conduct sensitivity analyses. These steps will help uncover uncertainties and biases and gauge their significance.

This paper was written while the author was a Wigner Fellow with the Energy Division of Oak Ridge National Laboratory, operated by Martin Marietta Energy Systems, Inc., under Contract No. DE-AC05-840R21400 with the U.S. Department of Energy. The author would like to thank Al Voelker, Bob Perlack, and Carl Petrich of ORNL for their encouragement and insights and an anonymous referee for helpful comments.

REFERENCES

Adelman, L., and Mumpower, J. 1978. A Critique of the Water Resources Assessment Methodology. Center for Research on Judgment and Policy, University of Colorado, Boulder.

Bacow, L. 1980. The Technical and Judgmental Dimensions of Impact Assessment. *Environmental Impact Assessment Review* 1(2):109–124.

Bakus, G. J., Stillwell, W. G., Latter, S. M., and Wallerstein, M. C. 1982. Decision Making: With Applications to Environment Management. *Environmental Management* 6(6):493–504.

Bisset, R. 1980. Methods for Environmental Impact Analysis: Recent Trends and Future Prospects. *Journal of Environmental Management* 11(1):27–43.

Brown, C. A., Quinn, R. J., and Hammond, K. R. 1980. *Scaling Impacts of Alternative Plans*. Denver, Colorado: US Bureau of Reclamation, Engineering and Research Center.

Brown, C. A. 1984. The Central Arizona Water Control Study: A Success Story for Multiobjective Planning and Public Involvement. *Water Resources Bulletin* 20(3):331–338.

Canter, L. W. 1979. *Environmental Impact Statements on Municipal Wastewater Programs*. Washington, D.C.: Information Resources Press.

Chankong, V., and Haimes, Y.Y. 1983. *Multiobjective Decision Making: Theory and Methodology*. Amsterdam:North-Holland.

Cohon, J. L. 1978. *Multiobjective Programming and Planning*. New York:Academic Press.

Collins, J. P., and Glysson, E. A. 1980. Multiattribute Utility Theory and Environmental Decisions. *Journal of the Environmental Engineering Division, American Society of Civil Engineers* 106, EE4:815–830.

Dee, N. *et al.* 1973. An Environmental Evaluation System for Water Resource Planning. *Water Resources Research* 9(3):523–535.

Dennis, R. L. et al. 1983. Integrating Technical and Value Issues in Air Quality Policy Formation: A Case Study. *Socioeconomic Planning Sciences* 17(3):95–108.

Dyer, J. S., and Sarin, R. K. 1979. Measurable Multiattribute Value Functions. *Operations Research* 27(4):810–822.

Elliot, M. L. 1981. Pulling the Pieces Together: Amalgamation in Environmental Impact Assessment. *Environmental Impact Assessment Review* 2(1):11–38.

Fischhoff, B. L., Goitein, B., and Shapira, Z. 1980. The Experienced Utility of Expected Utility Approaches. In N. T. Feather (ed.) *Expectancy, Incentive, and Action*. Hillsdale, NJ: Erlbaum Associates.

Fischhoff, B. L., Slovic, P., and Lichtenstein, S. 1979. Knowing What You Want: Measuring Labile Values. In T. Wallsten (eds.) *Cognitive Processes in Choice and Decision Behavior*. Hillsdale, NJ: Erlbaum Associates.

Gershon, M., and Duckstein, L. 1983. Multiobjective Approaches to River Basin Planning. *Journal of Water Resources Planning and Management* 109(1):13–28.

Goicoechea, A., Hansen, D., and Duckstein, L. 1982. *Introduction to Multiobjective Analysis with Engineering and Business Applications*. New York: John Wiley and Sons.

Gum, R. L., Roefs, R. G., and Kimball, D. B. 1976. Quantifying Social Goals: Development of a Weighting Methodology. *Water Resources Research* 12(4):617–622.

Hammond, K. R., Stewart, T. R., Brehmer, B., and Steinman, D. O. 1975. Social Judgment Theory. In M. F. Kaplan and S. Schwartz (eds.) *Human Judgment and Decision Processes*. New York: Academic Press.

Henderson, J. E. 1982. *Handbook of Environmental Quality Measurement and Assessment: Methods and Techniques. Instruction Report E-82-2*. Vicksburg, MS: Environmental Laboratory, U.S. Army Corps of Engineers Waterways Experiment Station.

Hobbs, B. F. 1979. *Analytical Multiobjective Decision Methods for Power Plant Siting: A Review of Theory and Applications. NUREG/CR-1687, BNL-NUREG-51204*. Upton, NY: Brookhaven National Laboratory.

Hobbs, B. F. 1980. A Comparison of Weighting Methods in Power Plant Siting. *Decision Sciences* 11(4):725–737.

Hobbs, B. F. 1985. Experiments in Multiple Criteria Decision Making and What We Can Learn From Them: An Example. In Y. Haimes and V. Chankong (eds.) *Decision Making with Multiple Objectives*. New York: Springer-Verlag.

Hobbs, B. F., and Rowe M. D. 1979.*A Comparison of Regional Screening Methods. NUREG/CR-1688, BNL-NUREG-51205* Upton, NY: Brookhaven National Laboratory.

Hobbs, B. F., Rowe, M. D., Pierce, B. L., and Meier, P. M. 1984. Comparisons of Methods for Evaluating Multiattributed Alternatives in Environmental Assessments: Results of the BNL-NRC Siting Methods Project. In S. Hart, G. Enk, J. Jordan, and P. Perreault (eds.) *Improving Impact Assessment: Increasing the Relevance and Utilization of Technical and Scientific Information*. Boulder, CO: Westview Press.

Hollick, M. 1981. The Role of Quantitative Decision-making Methods in Environmental Impact Assessment. *Journal of Environmental Management* 12(1):65–78.

Hyman, E. L. 1981. The Uses, Validity, and Reliability of Perceived Environmental Quality Indicators. *Social Indicators Research* 9(1):85–110.

Hyman, E., and Stiftel, B. 1980. Retrospective of a Research Program on the Theory and Practice of Environmental Impact Assessment. *Environmental Impact Assessment Review* 1(4):428–430.

Janssen, R. and Nijkamp, P. 1985. A Multiple Criteria Evaluation Typology of Environmental Management Problems. In Y. Haimes and V. Chankong (eds.) *Decision Making with Multiple Objectives*. New York: Springer-Verlag.

John, R. S., and Edwards, W. 1978. *Importance Weight Assessment for Additive, Riskless Preference Functions: A Review. Research Report 78-5*, Los Angeles Social Science Research Institute, University of Southern California.

John, R. S., Edwards, W., and Collins, L. 1980. *A Comparison of Importance Weights for Multiattribute Utility Analysis derived from Holistic, Indifference, Direct Subjective, and Rank Order Judgments. Research Report 80-4* Los Angeles: Social Science Research Institute, University of Southern California.

Keeney, R. L., and Robilliard, G. 1977. Assessing and Evaluating Environmental Impacts at Proposed Nuclear Power Plant Sites. *Journal of Environmental Economics and Management* 4:153–166.

Keeney, R.L., and Raiffa, H. 1976. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. New York: Wiley.

Krzysztofowicz, R. 1983. Strength of Preference and Risk Attitude in Utility Measurement. *Organizational Behavior and Human Performance* 31(1):88–113.

Lord, W.B., Deane, D.H., and Waterstone, M. 1979. *Commensuration in Federal Water Resources Planning: Problem Analysis and Research Appraisal*. Denver, Colorado: US Bureau of Reclamation, Engineering and Research Center.

Manheim, M. L. 1981. Ethical Issues in Environmental Impact Assessment. *Environmental Impact Assessment Review* 2(4):315–334.

McAllister, D. M. 1980. *Evaluation in Environmental Planning*. Cambridge, MA: MIT Press.

McClelland, G. H. 1979. *Equal versus Differential Weighting for Multiattribute Decisions: There Are No Free Lunches. Center Report 207*. Boulder, CO: Center for Research on Judgment and Policy, University of Colorado.

New York Department of Environmental Conservation 1980. Hazardous Waste: The Siting Board. *New York State Environment* 9 (December 18):3–6.

Newman, J. R. 1977. Differential Weighting in Multiattribute Utility Measurement: When It Should Not and When It Does Make a Difference. *Organizational Behavior and Human Performance* 20:312–325.

Nichols, R., and Hyman, E. 1982. Evaluation of Environmental Assessment Methods. *Journal of the Water Resources Planning and Management Division, American Society of Civil Engineers* 108, WR1:87–105.

Nisbett, R. E., and Wilson, T. D. 1977. Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review* 34:231–259.

O'Banion, K. 1980. Use of Value Functions in Environmental Decisions. *Environmental Management* 4(1):3–6.

Pierce, B. L., and Rowe, M. D. 1979. *Quantitative Nuclear Power Plant Siting Methods: A Review of Current Practice. BNL-NUREG-28115*. Upton, NY: Brookhaven National Laboratory.

Rowe, M. D., and Pierce, B. L. 1982a. Sensitivity of the Weighting Summation Decision Method to Incorrect Application. *Socio-Economic Planning Sciences* 16(4):173–177.

Rowe, M. D., and Pierce, B. L. 1982b. Some Tests of Analytical Multiobjective Decision-Making Methods. *Socio-Economic Planning Sciences* 16(3):133–144.

Saaty, T. L., Vargas, L. G., and Wendell, R. E. 1983. Assessing Attribute Weights by Ratios. *OMEGA* 11(1):9–12.

Sarin, R. K. 1980. Ranking of Multiattribute Alternatives with an Application to Coal Power Plant Siting. In *Multicriteria Decision Making, Theory and Application*. New York: Springer-Verlag.

Sarin, R. K., Dyer, J. S., and Nair, K. 1980. A Comparative Evaluation of Three Approaches for Preference Function Assessment. Presented at *The Institute of Management Sciences/Operations Research Society of America Joint National Meeting*, Washington, D C, May 4–7.

Schoemaker, P. J. H. 1981. Behavioral Issues in Multiattribute Utility Modeling and Decision Analysis. In J. N. Morse (ed.) *Organizations: Multiple Agents with Multiple Criteria*. New York: Springer-Verlag.

Schoemaker, P. J. H. 1982. The Expected Utility Model: Its Variants, Purposes, Evidence, and Limitations. *Journal of Economic Literature* 20(2):529–563.

Schoemaker, P. J. H., and Waid, C. C. 1982. An Experimental Comparison of Different Approaches to Determining Weights in Additive Utility Models. *Management Science* 28(2):182–196.

Shepard, R. 1964. On Subjectively Optimal Selections Among Multiattribute Alternatives. In M. Shelley and G. Bryan (eds.) *Human Judgments and Optimality*. New York: Wiley.

Stiftel, B., and Hyman, E. L. 1980. Assessment of Environmental Quality in a Democratic State. *The Environmental Professional* 2:306–314.

Susskind, L. E., and Dunlap, L. 1981. The Importance of Nonobjective Judgments in Environmental Impact Assessments. *Environmental Impact Assessment Review* 2(4):335–367.

Warner, M., and Preston, E. 1974. *A Review of Environmental Impact Methodologies, EPA-600/5-74-002*. Washington, DC: US Environmental Protection Agency.

Wu, J. S., and Hilger, H. 1984. Evaluation of EPA's Hazard Ranking System. *Journal of Environmental Engineering* 110(4):797–807.