Local and regional coherence utility assessment procedures

M.R. NOVICK; D.F. DEKEYREL
and
D.T. CHUANG
The University of Iowa

SUMMARY

Novick and Lindley (1978, 1979) have dealt with the use of utility functions for applications in education and have advocated the use of the standard gamble (von Neumann and Morgenstern, 1953) elicitation procedure with the addition of coherence checking using overspecification and a least squares fit. In this procedure utilities are inferred from probability judgements offered by the assessor. This paper describes local and regional coherence procedures which seek utility coherence in successive restricted domains of the parameter space as preludes to overall coherence checking. These procedures and some others are viewed as possible ways of avoiding anchoring and certainty effect biases found in earlier fixed probability methods, and presumably present in current fixed state procedures.

Keywords: UTILITY ASSESSMENT, COHERENCE

INTRODUCTION

Earlier approaches to utility assessment (Mosteller and Nogee, 1951, Schlaifer, 1959, 1971; Raiffa and Schlaifer, 1961; Keeney and Raiffa, 1976; and so on) have been based on the use of fixed probability (FP) assessment procedures in which utilities are elicited directly, through successive bisections of the parameter space. It has been suggested (Mosteller and Nogee, 1951) that such procedures are easier to use because subjects are more familiar with the quantity for which the utility function is desired than they are with probabilities, which they are required to state in the standard fixed state (SFS) procedure.

Although it was originally thought by psychologists that utility theory would prove useful as a *descriptive* model (Swalm, 1966, etc.), much criticism has recently been levied against its use in that capacity. As principal critics, Kahneman and Tversky (1978) have proposed an alternative descriptive model. The main basis for their criticism is that the phenomenon described by Tversky (1977) as the *certainty effect* results in preferences that violate the substitution axiom or expected utility hypothesis of utility theory. This axiom (hypothesis) states that preference order is invariant over probability mixtures and is formally equivalent to the assumption that there is no positive or negative utility for the act of gambling itself. Specifically, the *certainty effect* is the phenomenon that the utility of an outcome seems greater when it is certain that when it is uncertain. This effect can be observed when subjects are presented with a choice between a for-sure and a chance option, the choice appearing in the standard gamble, regional coherence, and local coherence assessment procedures to be described in this paper.

Utility theory as considered here is used as a *normative* model rather than as a *descriptive* model; however, it is still important to consider the certainty effect because Tversky (1977) has shown that even when subjects were told that their preferences violated utility theory, they were not inclined to change them (see also Kahneman and Tversky, 1972). This brings into question the reliability (coherence) and bias-free character of utililty assessment procedures obtained through both fixed state and fixed probability methods and the value of those procedures in helping decision makers be more coherent. However, it should be pointed out that the gambles studied by Kahneman and Tversky and those studied by Novick and Lindley were somewhat different and that the latter authors also provided incoherence resolution procedures.

In another paper, Tversky and Kahneman (1974) described several heuristics used by persons in assessing probabilities and the biases to which they could lead. Of particular interest is the *anchoring and adjustment heuristic*, which Spetzler and Staël von Holstein (1975) have shown can reduce the reliability with which the bisection technique used by fixed probability models elicit utilities. This heuristic is the phenomenon whereby the most readily available piece of information often forms an initial basis for formulating responses from which subsequent responses are then adjusted. Since adjustments from this basis are often insufficient, a central bias results. According to Slovic (1972, the *anchoring and adjustment heuristic* is a natural strategy for easing the strain of integrating information. The anchor serves as a register in which one stores first impressions or the results of earlier calculations. Why adjustments from the anchor are usually insufficient, though, is unclear. Slovic advances two hypotheses to explain the insufficient adjustment. First, people may stop adjusting too soon because they tire of the mental effort involved in adjusting. Alternatively, the anchor may take on a special salience, thus causing people to feel that there is less risk in making estimates close to it than in making estimates that deviate far from it. According to Spetzler and Staël von Holstein (1975), experimentation has shown that subjects tend to produce a central bias when, in the fixed probability bisection method, they are asked first for the median for an uncertain quantity and then for the quartiles.

Later, in reviewing the role of man-machine systems in decision analysis, Slovic, Fischhoff, and Lichtenstein (1977) suggested that human factors such as the ways in which variations in instructions or informational displays affect people's performance are important and should be studied in more detail. Questions of complexity and representativeness of material seem to have substantial effect on assessors responses (Fischhoof, Slovic and Lichtenstein, 1977; Vlek, 1973). The study of such factors might lead to an assessment procedure that minimizes the judgemental basis and heuristics described earlier. This position was strengthned by the discussion of Fischhoof, Slovic and Lichtenstein (1979). A consideration of these ideas promoted the development of a new format introduced later in this paper.

Extensive previous work in this area has raised more questions concerning bias and coherence than it has provided answers. An apparently pessimistic mood prevails, not inappropriately, given the importance of the questions that have been raised (Hogarth, 1975; Slovic, 1975; and Fischhoff, Slovic and Lichtenstein, 1979). Nevertheless, the very extensiveness of this research must itself imply a high assessment for the product of the probability of resolving these difficulties and the value of this outcome. The position taken here is that bias and incoherence can be reduced if (1) elicitations are carefully fashioned in a Computer-Assisted Data Analysis (CADA) environment (Novick, Hamer, Libby, Chen and Woodworth, 1980), (2) assessors are aided in resolving incoherence, and (3) if the assessments concern states and actions that are meaningful and important to the assessor at the time the assessment is made.

Consider a variable θ and the utility function $U(\theta)$ for which assessment is required. In most applications θ will be a real variable, such as grade point average (GPA), but this is not necessary. Although the contrary assumption is sometimes made, it seems sensible to us to demand that a utility function be bounded and increasing.

There are two standard approaches to assessing a utility function: fixed probability and fixed state. In the former, the subject is presented with a gamble on two values, or states, θ_1 and θ_2 with a fixed probability-p, say, for θ_1 and $1 - p$ for θ_2 - and is required to choose an intermediate state θ_3 such that he/she is indifferent with respect to the gamble and θ_3 for sure. In applications, typically $p = 1/2$ because this gamble is easiest for assessors (subjects) to understand.

In the fixed-state method, the states θ_1, θ_2, and θ_3 are fixed, θ_3 still being intermediate between θ_1 and θ_2. The subject is required to state a probability,

p, such that he/she is indifferent between θ_3 for sure and the following gamble: θ_1 with probability p and θ_2 with probability 1-p. If θ_1 and θ_2 have utilities of 1 and 0, respectively, the gamble has expected utility p, the indifference probability assigned to θ_3.

In the fixed-state method, let us suppose that a number of states, θ_0, $\theta_1,\ldots,\theta_{N+1}$, are selected. We shall further suppose that these states are ordered in the sense that θ_j is preferred to θ_i whenever $j > i$; in particular θ_{N+1} is the best and θ_0 the worst state. Then the utility function $U(\theta)$ will be strictly increasing.

Without loss of generality, the utility for θ_{N+1} can be assigned the value 1 and that for θ_0 can be assigned the value 0, thus placing bounds on the utility values to be assigned to the various states. We must then find N such values: $U(\theta_1)$, $U(\theta_2),\ldots,U(\Theta_N)$. We first consider adjacent gambles, that is, a situation in which the subject is asked to compare the sure outcome $\theta_n(1 \leq n \leq N)$ against a gamble with possible outcomes θ_{n-1} and θ_{n+1}, representing, because of the ordering of the states, situations respectively worse and better than θ_n. Specifically, after a brief review of the meaning of probability, the subject is asked to state the probability p_n for θ_{n+1}, and consequently 1 - p_n for θ_{n-1}, that makes him/her indifferent with respect to the gamble and θ_n for sure. Writing $U(\theta_n) = u_n$ (so that $u_0 = 0$, $u_{N+1} = 1$) and equating the expected utilities for the two situations gives us

$$u_n = p_n u_{n+1} + (1 - p_n)u_{n-1}.$$

If this done for all n, $1 \leq n \leq N$, we have N equations in N unknowns and aside from exceptional cases, the utilities are uniquely determined. The solution is

$$u_{n+1} = G_n/G_N$$

for $0 \leq n \leq N$, where

$$G_n = \Sigma_{i=0}^{n} F_i,\ F_n = \Pi_{i=0}^{n} f_i, \text{ and } f_i = (1 - p_i)/p_i.$$

Suppose a subject has responded to the N question previously considered and, from the answers given, his/her utilities u_1, $u_2,\ldots,u_N$ have been determined. Suppose also that he/she is asked to consider a gamble that will yield either θ_{n+2} or θ_{n-2}, against v_n for sure. Then the probability q_n, associated with θ_{n+2}, satisfies

$$u_n = q_n u_{n+2} + (1\text{-}q_n)u_{n-2}.$$

For the fixed state standard gambles procedure the suggestion offered by Novick and Lindley is that to exploit coherence fully, we must ask for more probability assessments than are needed to calculate the utilities and then compare them for coherence. The idea of requiring the experimenter to give more than the minimum number of judgments in fitting a personal probability distribution has been used by Pratt, Raiffa and Schlaifer (1965) and has been exploited systematically both for the assessment of probabilities and utilities in the development of the Computer Assisted Data Analysis Monitor (CADA) (Novick, 1973, 1975). In the context of utility assessments, the idea has been used by Becker, DeGroot, and Marschak (1963) with fixed probability assessments, and we shall discuss this presently.

Experience shows us that assessors are almost always incoherent but readily attempt to resolve their incoherences when these are brought to their attention (cf. MacCrimmon, 1965). It may, however, be true that one kind of gamble (e.g., adjacent gambles) may introduce one kind of systematic bias and another kind (e.g., extreme gambles) may introduce a second kind of bias. Therefore, rather than just asking the subject to revise some of his/her assessments, Novick and Lindley (1979) suggest assisting the subject by providing a least squares fit in the log-odds metric for the N undetermined utility values. A computer program has been written to carry out the interrogation of the assessor and to perform the least squares fit and is available on the CADA Monitor (Novick, et. al., 1980).

In any comparison of fixed state with fixed probability assessments, the role of coherence seems to us to play a dominant role. Although subjects often prefer fixed probability assessments, especially when the probability is 1/2, exploiting coherence in this context is harder than with the fixed state procedure. For example, suppose, as usual, that a subject is asked for the certainty equivalent of a gamble, at even odds, on the best (θ_{N+1}) and worst (θ_0) states. Let his/her stated value be θ_m, say, having $u(\theta_m) = 1/2$. The subject is then asked for the certainty equivalents for even-odds gambles on (θ_0, θ_m) and (θ_m, θ_{N+1}). If these values are denoted by θ_1 and θ_2, respectively, then the utilities of θ_1 and θ_2 are $u(\theta_1) = 1/4$ and $u(\theta_2) = 3/4$. Finally, he/she is asked to consider an even-odds gamble on θ_1 and θ_2. But it is rather transparent that for coherence the result must be θ_m, so that the four judgments can scarcely be considered independent. In this field (as in other measurement fields) obtaining independent repetitions of the same assessment is hardly ever possible, thus the emphasis ought to be on independent assessments of related quantities. This, we feel, is more nearly achieved with the fixed-state assessments. The above discussion is taken with some condensation from Novick and Lindley (1979).

The question that must now be addressed is whether the incoherence resolution of the least squares method described in SFS above avoids the certainty and anchoring effects or whether better methods can be found. The remainder of this paper will be devoted to describing a refinement in the least squares SFS procedure and in describing three new procedures that more directly address these biasing effects.

A word concerning ease of response may be in order. Mosteller was certainly correct in saying that FP is easier than FS, and, indeed, without interactive conversational computing facilities an FS assessment procedure may well be unbearably difficult. With conversational computing, however, an FS procedures is bearable and there is no reason to believe the easier method is more bias free. Indeed, the contrary could be true.

In the current version of the SFS procedure on CADA, subjects are given situations consisting of a for-sure and a chance option on grade point averages in the range 0-4 and are asked for the probabilities that make them indifferent with respect to the two options in each situation (i.e., their indifference probabilities). The indifference probabilities for the fixed state gambles are elicited using one of two formats for presenting the gambles. Format two request a direct magnitude estimation as illustrated earlier. Format one asks for preferences for gambles or sure things for p values .1, .9, .2, .8, etc., or .9, .1, .8, .2, etc., with zeroing in on the indifferent point.

FORMAT ONE

3.0 p chance	indifferent = 0
2.5 for sure	for sure = 1
2.0 1-p chance	chance = 2
	restart = 3

Which would you prefer if p were .XX ?______. (This question was repeated using the following p values .1, .9, .2, .8, ... until p was found to be between .5 and .6, say. Then the questioning procedure used p values of .52, .58, .54, ... until the subject's indifference p had been determined).

FORMAT TWO

for sure	gamble with prob p	 with prob 1-p	p that makes you indifferent
2.5	3.0	2.0	?______.

TABLE 1 *Formats for the SFS utility assessment procedure*

Format two, the direct probability assessment format is the one used by Novick and Lindley, (1979). Format one, the *ends-in procedure,* has been advanced as a method for avoiding anchoring. Since indifference points are typically between .2 and .8 any initial anchor (.1 or .9) is erased before any careful judgment must be made. Also, the starting values alternate between .1 and .9 thus avoiding any constant ordering effect. It is our as-yet-unsubstantiated belief that this format is both easier and less subject to anchoring than format two. This format is now used with several assessment procedures.

In order to avoid the documented biases of the certainty effect in utility assessment, a new procedure has been considered: the paired binary gambles (PBG) procedure. This procedure is illustrated in Table 2 bellow. The paired gambles in the table can be abbreviated as (1.5 3.0, 2.0 2.5).

Paired Binary Gambles

SITUATION 1	**SITUATION**	**2**
3.0 p	2.5	p
1.5 1-p	2.0	1-p

TABLE 2 *PBG procedure*

The ends-in format is used to elicit the probability that will make the subject indifferent with respect to the two situations (gambles). A least squares fit of the indifference probabilities can then be made and subjects can proceed as in the SFS procedure.

Although the PBG procedure is considered here as a fixed state procedure, it has previously been used in a fixed probability paradigm (Kneppreth, Gustafson, Leifer, & Johnson, 1974). Suppes and Walsh (1959) have considered such gambles strictly in the sense of determining preferences between the two situations, without eliciting either indifference probabilities or equivalance points.

The obvious hope is that the PBG procedure will avoid the certainty effect because the comparison is between two sets of gambles, and thus does not involve the for-sure option. We have used PBG in some informal assessments but have not yet been convinced of its usefulness. First, it is difficult even for experienced subjects. Fatigue and boredom are definite problems. We are not sure that there is not a bias in that one situation always compares two adjacent states while the other always describes two states twice removed. We have not discarded this procedure, but we feel that refinements may be necessary if it is to be useful.

Next we define the Regional Coherence (RC procedure). In the RC procedure, indifference probabilities are elicited separately for two SFS gambles using the ends-in format. Subjects are then presented with a table showing the initial gambles (situations 1 and 2 with their indifference probabilities) and two additional gambles (situations 3 and 4). They are told that their initial responses imply certain specific indifference probabilities for the two new gambles. Table 3 illustrates the latter part of this procedure.

	Situations			
	1	2	3	4
p-chance	1.50	2.00	2.00	2.00
for sure	1.00	1.50	1.00	1.50
1-p chance	0.50	1.00	0.50	0.50
	$p=.53$	$p=.58$	$p=.40$	$p=.75$

TABLE 3 *RC Procedure*

Assessors are then given the opportunity to change the indifference probabilities, two at a time, until they are indifferent in all four situations. They choose the two situations for which they wish to change the indifference probabilities and the magnitude estimation format is then implemented to generate the revised probabilities.

The final procedure is called the local coherence (LC) procedure. This procedure presents subjects with two types of hypothetical choice situations: (1) a for-sure and a chance option (the standard gamble) and (2) two chance options. The ends-in format is used to elicit an indifference probability for the first situation, after which the subject is told that that response implies that he/she should be indifferent with respect to the two options in situation 2. Note that the subject only specifies the indifference probability for the standard gamble. Table 4 below illustrates this procedure. The probabilities for the second situation are uniquely determined by that specification.

SITUATION 1		SITUATION 2		
		option one		option two
4.00	.75 chance	.19	4.00	---
3.00	for sure	---	3.00	.25
1.00	.25 chance	.81	1.00	.75

TABLE 4 *LC Procedure*

If the subject is not indifferent in both situations, he/she modifies the situation 1 indifference probability and then is again presented with a table similar to Table 4 above. This continues until subject is indifferent with respect to the two situations.

In choosing a fixed state assessment procedure we are free to select

(1) A response format
  a. ends-in
  b. direct specification
(2) A comparison format
  a. standard fixed state
  b. paired binary gambles
  c. regional coherence
  d. local coherence
(3) Overall coherence checking by least squares
  a. yes
  b. no

The temptation for a person trained in both psychology and statistics to undertake the experimental comparison using some subset of a 2 by 4 by 2 factorial design is overpowering. Indeed the tooling-up for this experiment has begun including a further comparison with the fixed probability method and an investigation of comparative bias for central and extreme values of θ. For a Bayesian statistician, yielding to this temptation leads to a compulsion to state a prior distribution. In the absence of a precisely stated model this is not possible, but it is possible to state some general beliefs. I shall now do this and also invite you to attend the Psychometric Society meetings in May of 1980 where I shall report on the results of these experiments.

First, I believe that SFS with overall LSQ coherence checking will prove to be good but not best. Subjects find it hard to make unaided adjustments. As a result, incoherence will remain high, but overall fits will be tolerably good ($p = .7$). However, we are working on improvements that could make this procedure more attractive. I believe that the ends-in format will be preferred over direct magnitude estimation and will reduce the anchoring effect ($p = .8$). This may not hold for very experienced assessors who may find it tedious.

I believe that PBG will be unpopular and ineffective unless we find some simplification ($p = .9$). At present it is difficult and fatigueing and responses tend to be less than carefully considered.

I believe that regional and local coherence will both be useful and both will largely eliminate anchoring and adjustment biases ($p = .8$). I believe the regional coherence will be preferred by inexperienced users ($p = .6$) and local coherence by experienced users ($p = .7$, but perhaps only professional statisticians). Local Coherence provides a display of the large effect on extreme comparisons of minor adjustment in non-extreme comparisons. It is a powerful tool for locating the most desirable point in the probability range $p \pm .025$. It is unclear to me whether overall least squares overfitting will be useful in conjunction with the RC or LC procedures, but my prior probability is .6 that it is.

Finally, for certain points I have high personal probability. Utility elicitation procedures can currently be conducted with accuracy and ease on CADA. Indeed, they can be conducted with sufficient ease and potentially with sufficient freedom from bias as to make applications of utility theory to education entirely feasible when the assessor is confronted with a specific problem of interest and importance, and when that problem is presented clearly and unambiguously.

ACKNOWLEDGEMENT

Research supported by the Office of Naval Research under contract N00014-77-C-0428.

REFERENCES

DEGROOT, M.H. (1970). *Optimal statistical decisions.* New York: McGraw Hill.

— (1975). *Probability and statistics.* Reading, Mass: Addison-Wesley.

FISCHHOFF, B., SLOVIC, P., & LICHTENSTEIN, S. (1977). Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Tech. Rep.* PTR--1042-77-8, University of Iowa.

— (1980). Knowing what you want: Measuring labile values. Appearing in *Cognitive Processes in Choice and Decision Behavior.* (T. Wallston, ed.) Hillsdale, New Jersey: Laurence Erlbaum Associates, (In press).

GROSS, A.L., & SU, W.H. (1975). Defining a "fair" or "unbiased" selection model: A question of utilities. *J. Appl. Psychol.* **60**, 345-351.

HOGARTH, R.M. (1975). Cognitive processes and the assessment of subjective probability distributions. *J. Amer. Statist. Assoc.* **20**, 271-289.

KAHNEMAN, D., & TVERSKY, A. (1978). Prospect theory: An analysis of decision under risk. *Econometrica.* **47**, 263-291.

— (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology,* **3**, 430-454.

KEENEY, R.L., & RAIFFA, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs.* New York: Wiley.

KNEPPRETH, N.P., GUSTAFSON, D.H. LEIFER, R.P. & JOHNSON, E.M. (1974). Techniques for the assessment of worth. *U.S. Army research institute for the behavioral and social sciences. Tech. Rep.* **254**, 40-53.

LEHMANN, E.L. (1959). *Testing Statistical Hypothesis.* New York: Wiley.

MOSTELLER, F., & NOGEE, P. (1951). An experimental measurement of utility. *Political Economy,* **59**, 371-404.

NOVICK, M.R. (1973). High School attainment: an example of a computer-assisted Bayesian approach to data analisis. *Internat. Statist. Rev.* **41**, 269-271.

— (1975). A cource in Bayesian Statistics. *Amer. Statist.* **29** , 94-97.

NOVICK, M.R. & ISAACS, G.L. (1978). Manual for the computer-assisted data analysis (CADA) Monitor, Iowa City: The University of Iowa.

NOVICK, M.R., ISAACS, G.L., HAMER, R., CHEN, J., CHUANG, D., WOODWORTH, G., MOLENAAR, I., LEWIS, C. & LIBBY, D. (1980). *Manual for The Computer-assisted data analysis monitor*, Iowa City: The University of Iowa

NOVICK, M.R. & JACKSON, P.H. (1974). *Statistical methods for educational and psychological research.* New York: McGraw Hill.

NOVICK, M.R., & LINDLEY, D.V. (1979). Fixed state assessment of utility functions. *J. Amer. Statist. Assoc.* **74**, 306-311.

— (1978). The use of more realistic utility functions in educational applications. *J. Educational Measurement,* 181-192.

NOVICK, M.R. & PETERSEN, N.S. (1976). Towards equalizing educational and employment opportunity. *J. Educational Measurement,* **13**, 77-88.

PETERSEN, N.S. (1976). An expected utility model for "optimal" selection. *J. Educational Statistics.* **1**, 333-358.

PETERSEN, N.S. & NOVICK, M.R. (1976). An evaluation of some models for culture-fair selection. *J. Educational Measurement,* **13**, 3-31.

RAIFFA, H., & SCHLAIFER, R. (1961). *Applied statistical decision theory.* Boston: Harvard University.

SCHLAIFER, R. (1959). *Probability and statistics for business decisions: An introduction to managerial economics under uncertainty.* New York: McGraw Hill.

SLOVIC, P. (1975). Choice between equally valued alternatives. *Experimental Psychol.* **3**, 280-287.

— (1972). From Shakespeare to Simon: Speculation -and some evidence- about man's ability to process information. *Oregon Research Institute Research Bulletin,* **12**, 1-29.

SLOVIC, P., FISCHHOFF, B., & LICHTENSTEIN S. (1977). Behavioral decision theory. *Annual Rev. Psychol.,* **28**, 1-39.

SPETZLER, C.S. & STAEL VON HOLSTEIN, C.A.S. (1975). Probability encoding decision analysis. *Management Science,* **22**, 340-358.

SUPPES, P., & WALSH, K. (1959). A non-linear model for the experimental measurement of utility. *Behavioral Science,* **4**, 204-211.

SWALM, R.O. (1966). Utility theory--insights into risk taking. *Harvard Business Rev.* **44**, 123-136

TVERSKY, A. (1977). On the elicitation of preferences: Descriptive and prescriptive considerations. In *Conflicting objectives in decisions,* (D.E. Bell, R.L. Keeney and H. Raiffa eds.) 209-222. New York: Wiley.

TVERSKY, A. & KAHNEMAN, D. (1974). Judgment under uncertainty: Heuristics and biases *Science,* **185**, 1124-1131.

VAN DER LINDEN, W.J. & MELLENBERGH, G.J. (1977). Optimal cutting scores using a linear loss function. *Appl. Psychol. Measurement,* **1**, 593-599.

VLECK, C.A.J. (1973). Coherence of human judgment in a limited probabilistic environment. *Organizational Behavior and Human Performance,* **9**, 460-481.

VON NEUMANN, J. & MORGENSTERN, O. (1953). *Theory of games and economic behavior.* Princeton: University Press.