PROCEDURES FOR AGGREGATING CITIZEN PREFERENCES IN THE CONTEXT OF THE NUCLEAR WASTE MANAGEMENT PROBLEM

Final Report

October 1978

MASTER

Prepared for:

Lawrence Livermore Laboratory
University of California

Go #8942403

SRI International
333 Ravenswood Avenue
Menlo Park, California 94025
(415) 326-6200
Cable: SRI INTL MNP
TWX: 910-373-1246

DISCLAIMER

This book was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of its employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for its accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

**PROCEDURES FOR
AGGREGATING CITIZEN
PREFERENCES IN THE CONTEXT
OF THE NUCLEAR WASTE
MANAGEMENT PROBLEM**

Final Report

October 1978

**By: Horance W. Brock
Gerald L. Sauer**

**Prepared for:
Lawrence Livermore Laboratory
University of California**

SRI Project 7093

CONTENTS

LIST OF TABLES	iv
I INTRODUCTION	1
I.A Perspective	2
II TWO APPROACHES TO SOCIAL DECISION ANALYSIS	5
II.A Initial Desiderata for the Analytical Procedure	6
II.B Implications of the Desiderata	6
II.C Insensitivity of V and M to "Strength" of Preferences	7
II.D Mechanics of S and P	8
III A JUSTIFICATION FOR THE USE OF EITHER S OR P IN TERMS OF FOUR FUNDAMENTAL NORMATIVE CRITERIA	11
III.A Introduction	11
III.B Four Fundamental Criteria	11
III.C "Needs-Respecting" Social Decisionmaking	12
III.C.1 Allocation of Utility in Direct Proportion to Relative Needs: P Aggregation	13
III.C.2 Allocation in Approximate Accord with Relative Needs: S Aggregation	18
III.D Impartial Social Decisionmaking	20
III.D.1 P Aggregation and the Concept of Impartial Social Decisionmaking	21
III.D.2 S Aggregation and the Concept of Impartial Social Decisionmaking	22
III.E Individual Rationality of S and P	23
III.E.1 Individual Rationality of S	23
III.E.1.a Interpersonal Comparison Coefficients	25
III.E.2 Individual Rationality of P	28
III.E.2.a Bargaining Model	28
III.E.2.b Zeuthen's Principle	29
III.E.2.c Derivation of P from Zeuthen's Principle	30

III.F	Collective Rationality of S and P	31
III.F.1	Collective Rationality of S	31
III.F.2	Collective Rationality of P	34
III.G	Conclusion	36
IV	THE APPLICATION OF S IN THE MEASUREMENT OF THE PERFORMANCE OF NUCLEAR WASTE MANAGEMENT SYSTEMS	37
IV.A	The Analytical Problem Simplified	38
IV.B	Present Versus Future	43
IV.C	Voluntary Versus Involuntary	44
V	INFERRING SOCIAL TRADEOFF PARAMETERS FROM EXPERT AND LAYMEN JUDGMENTS	45
V.A	Data	46
V.B	Averaging	47
V.C	Prior Information	48
V.C.1	Statistical Quick Death Versus Dollars	48
V.C.2	Using the Expert Prior to Screen Response on the Dollar/Statistical Death Tradeoff	51
V.C.3	Other Health Effects Versus Statistical Quick Deaths	51
V.C.4	Occupational Versus Nonoccupational and Future Versus Present Health Effects	52
V.D	Likelihood of the Responses of the Advisors	53
V.E	Inference Results	55
V.E.1	Posterior Distributions	55
V.E.2	Posterior Distributions for f_{1000}^{25} and $f_{10,000}^{25}$	56
V.F	Numerical Results	56
REFERENCES		58
ATTACHMENTS		
1	EXAMPLES OF POSSIBLE HEALTH EFFECTS OF EXPOSURE TO RADIATION	60
2	JUDGMENTS OF THIRTEEN NATIONAL ADVISORS ON SOCIAL TRADEOFFS CONCERNING HEALTH EFFECTS AND PREVENTATIVE EXPENDITURES	62
3	SMALL-RISK VALUES OF LIFE RESULTING FROM A SENSITIVITY ANALYSIS OF THE PREFERENCES OF A HYPOTHETICAL 25-YEAR-OLD MALE	67

TABLES

1	Present Inventory and Generation Rate of Radioactive Wastes from Either Spent Fuel or Its Equivalent in High-Level Wastes After Processing	2
2	Parameters (n'' and s'') of the Posterior Inverted-Gamma-1 Distributions	55
3	Summary of Hypothetical Inference Results	57
2-1	Importance of Health Effects Relative to Quick Deaths	63
2-2	Dollar Values for Preventing a Quick Death 30 Years from Now as a Result of a Nonoccupational Risk	64
2-3	Relative Importance of Nonoccupational Versus Occupational Health Effects	65
2-4	Relative Importance of Fatal Cancers Occurring at Different Times	66

I INTRODUCTION

One of the most difficult aspects of the analysis of nuclear waste management alternatives is the need to arrive at a decision that takes into account the legitimately differing preferences of different members of society. Because of differences in geographical location, lifestyle, or cultural background, different people will have different preferences regarding the safety-versus-economics tradeoffs that will inevitably be embodied in a national waste management policy. Nonetheless, some policy must ultimately be adopted by government; and it is to be hoped that this policy will in some sense reflect the differing preferences of citizens, or will at least represent a "fair compromise" among their alternative wishes.

The problem of identifying a "socially optimal" policy against a background of differing values is sometimes called the problem of "collective choice." At the heart of this problem is the matter of how to aggregate individual values (preferences) into "collective" or "social" values. The purpose of the present paper is to provide an introduction to the theory of social choice and related disciplines, and to relate this theory to the concrete problem of nuclear waste management.

In Section I of this report, we provide an overview of the problem. In Section II, we informally identify two candidate preference aggregation procedures that can be used to identify a socially optimal waste management policy. In Section III, we present a somewhat lengthy defense of the use of these two aggregation procedures. Each is shown to be compatible with four intuitively appealing criteria of collective decision-making. In Section IV we discuss the application of one of the procedures to the evaluation of waste management alternatives. In Section V we address the problem of inferring evaluation parameters from expert and laypersons' judgments.

I.A. Perspective

Table 1 shows the current inventory and rate of generation of high-level nuclear wastes in the United States.

Table 1

PRESENT INVENTORY AND GENERATION RATE OF RADIOACTIVE WASTES
FROM EITHER SPENT FUEL OR ITS EQUIVALENT
IN HIGH-LEVEL WASTES AFTER PROCESSING

<u>Sources</u>	<u>Cubic Meters</u>	
	<u>Spent Fuel</u>	<u>High-Level Waste</u>
Military stock	280,000 (mixed)*	
Commercial stock	3,500	2,300
Annual rate of increase		
Present industry (1 yr)	2,200	or 512
One reactor for life (30 yr)	1,050	or 240

* Mixed spent fuel and high-level waste.

Source: Lawrence Livermore Laboratory, Energy and Technology Review, University of California, October 1977

The waste is stored in temporary facilities, with most of the commercial waste being held on reactor sites. The problem confronting the nation is what to do with the waste over the longer term.

Present arrangements have produced no significant health problems so far; but safer, economically feasible alternatives appear to be available. Failure to agree on an alternative (after proper testing) means that the country will be faced with higher than necessary health risks and that commercial reactors could be pointlessly shut down if they were to run out of temporary storage.

Congress has given the Nuclear Regulatory Commission (NRC) the responsibility of solving the waste storage problem. It is a difficult

assignment because of the complex and uncertain technical, ethical, and political issues involved. (Examples of these issues are: Can wastes be stored underground in an affordable manner so that even under the worst geological scenarios they have only a very small chance of coming into contact with man? How can higher electricity costs to present consumers be compared with the risks of nuclear contamination to future generations? Lastly, what kind of policy would be acceptable to political representatives of citizens living near waste transportation routes or storage depots?)

In carrying out its mission, the NRC is continuously involved in exploring the potential effects of different technical alternatives, evaluating the alternatives in terms of their social desirability, and trying to achieve in the political arena the policy the NRC deems to be most in the public interest. The range of policies being considered at any time will change as more information comes to light, particularly information regarding technical and political constraints.

Analysis has a role to play in aiding the NRC in each of the above three activities. Systems analysis and "cost-benefit" analysis can be used to explore the effects of policy alternatives on the citizenry at large. Social choice theory can be utilized to make unavoidable ethical comparisons and to rank policy alternatives in terms of their overall desirability. Lastly, concepts from game and bargaining theory can be used, along with legal analysis, in thinking about the federal, state, and local political gauntlet that any successful policy must go through.

Too often, in our opinion, analysis has been limited to the first activity--the linking of policy with effects. Unfortunately, the results of such investigations, although useful, are often incomplete or unrealistic in the eyes of top-level decisionmakers, who must also grapple with complex ethical issues and political constraints. In the case of nuclear waste management, we feel that ethical issues ought to be addressed ahead of time, so that policymakers will not have to make complex tradeoffs, under pressure, with little guidance beforehand.

The rest of this paper is devoted to developing analytical procedures that incorporate ethical considerations and that can be applied to making the safety/cost tradeoffs inherent in nuclear waste management. More specifically, two procedures will be defined by which public policy decisions can be analyzed in general. Unlike many cost-benefit approaches, the proposed methods confront squarely the ethical and distributive problems that inevitably arise in public policymaking.

An attractive feature of the procedures, worth mentioning at the outset, is that they can be used to facilitate public participation in administrative deliberations.

II TWO APPROACHES TO SOCIAL DECISION ANALYSIS

Specifically, the analytical problem we are addressing is how to select from a set, Z , of policy alternatives, the one, $F(Z)$, that is most "socially desirable."

For the benefit of the nontechnical reader, our approach to answering this question will be informal. We will introduce and define in English several principles that are frequently appealed to in ethical decisionmaking. We will then narrow the candidate procedures down to two: procedures S and P. Procedure S entails maximizing total utility, i.e., providing "the greatest good for the greatest number of people." Procedure P entails the simulation of fair bargaining among the citizenry. S here refers to "sum," whereas P refers to "product" (for reasons made clear further on). Our discussion will bring out the different informational requirements: S requires interpersonal utility comparison coefficients, and P requires the identification of a "fair no agreement" outcome.

Next, we will introduce four additional normative criteria and show how both S and P are compatible with these criteria. This is a remarkable fact, and a compelling argument for using either S or P--especially because we know of no other procedures that are consistent with all the normative desiderata. However, in establishing this result we shall demonstrate that S and P are compatible with rather different versions of the four criteria. This should come as no surprise, of course. S and P are not at all identical, and in practice, the use of S will identify as socially optimal a policy that would not be so identified by the use of P.

II.A. Initial Desiderata for the Analytical Procedure

Any analytical procedure, it seems to us, must satisfy the following criteria:

- (a) Citizen sovereignty--The procedure should reflect citizens' individual preferences as much as possible.
- (b) Cognizance of uncertainty--The procedure should be able to explicitly take into account uncertainty, including citizens' preferences toward risky outcomes.
- (c) Anonymity--The procedure should not discriminate among citizens on grounds other than their respective needs or preferences.

II.B. Implications of the Desiderata

Desiderata (a) requires that the procedure should be based on the "utility" that citizens derive from the effects of policy alternatives. Symbolically, this means that the procedure should depend on Z through U , a vector of utility functions of representative citizens.

$$F = F[U(Z)] \quad . \quad (1)$$

(A citizen's utility function assigns a number or ranking to a policy effect, where more preferred effects receive a high number or ranking. Utility functions can be determined by an explicit elicitation procedure.)

A procedure, F , that derives social preferences from citizens' preferences, is called a Preference Aggregation Procedure.

Desiderata (b) requires that the U contain information on how citizens rank gambles among the potential effects (as opposed to effects that are certain to occur). This kind of information is contained in a special type of utility function--the von Neuman-Morgenstern (VNM) utility function. VNM utility functions exist under mild assumptions about rational decisionmaking [see Marshak (1950) as well as subsection III.F.1 below]. Such a function can be obtained by asking a citizen to choose between facing a given set of effects that are certain to occur, or facing a simple gamble between two other sets of effects. (The method is illustrated in subsection III.E.1.)

The use of VNM utility functions greatly simplifies the calculus of rational decisionmaking. Furthermore, such functions contain information on the "strength" of an individual's preferences similar to the way temperature readings contain information on heat. (The zero point and scale of the function is arbitrary, as in Celsius vs. Fahrenheit, but the function can be used to compare an individual's welfare gains just as temperature readings can be used to compare temperature changes.) Thus desiderata (a) is another reason for using VNM utility functions because they contain additional information (beyond rank order) on citizens' preferences (Harsanyi: 1977).

Anonymity, desideratum (c), means that it does not matter who has what preferences or needs. Symbolically, F must be symmetric in its VNM utility arguments:

$$F[u_1(z), u_2(z), \dots, u_i(z), \dots, u_j(z), \dots, u_n(z)] = F[u_1(z), u_2(z), \dots, u_j(z), \dots, u_i(z), \dots, u_n(z)] \text{ for } i, j = 1, 2, \dots, n \quad (2)$$

Many preference aggregation procedures satisfy (1) and (2); among these are the following four rather well-known procedures:

(V) Majority voting.

(M) "Maxi-Min" [making the worst citizen or group of citizens as well-off as possible-- $\text{Max}_Z \text{Min}_i u_i(z)$].

(S) Maximizing total utility [$\text{Max}_Z \sum_i u_i(z)$].

(P) Simulating fair bargaining [$\text{Max}_Z \prod_i u_i(z)$].

II.C. Insensitivity of V and M to "Strength" of Preferences

V and M do not require the degree of interpersonal comparison that S does. Nor do they require the definition of a "no-agreement" outcome as does P. Both V and M, however, can be insensitive to the strength of individual's preferences, as illustrated by the following example.

Suppose there are two alternatives, z_1 and z_2 . Suppose z_1 provides 5 cents in benefits to each Alaskan. Suppose that z_2 , on the other hand,

provides only 4 cents to Alaskans, except for Fortuna, who would benefit by \$10 million. Lastly, suppose everyone agrees that Fortuna is not the worst off Alaskan. Then both V and M would pick z_1 over z_2 . This is true no matter how favorable we make z_2 to Fortuna. Thus both V and M are ruled out by the following desideratum:

- (d) Sensitivity to strength of preferences--The procedure should be sensitive to the strength of individual's preferences.

[We state (d) in loose terms because a precise definition would require the introduction of a complex technical notation.]

S and P, on the other hand, do satisfy (d) because they are based on additive and multiplicative functions, respectively, which are continuous and increasing in each citizen's utility.

Although in our opinion (d) is sufficient to rule out V and M, we will introduce in the next section four new desiderata and see that both S and P satisfy these criteria also. Moreover, we are not aware of any demonstrations showing that either V or M satisfy these additional principles.

Before introducing the additional criteria, it will be helpful to discuss the operational use of our candidates S and P.

II.D. Mechanics of S and P

The first two steps in applying S and P are the same.

Step 1 S & P--By elicitation or other means, n VNM utility functions are obtained, where the i th one, u_i , represents the typical preferences of a fraction f_i of the citizenry. The preferences concern the effects of policy alternatives (e.g., health effects and storage costs for nuclear waste management), as opposed to the alternatives themselves, which may be too technical and complicated for the average layman to evaluate on his own. We remind the reader that both the zero point and the scale of VNM functions are arbitrary.

Step 2 S & P--Models are used to calculate the expected utility to each of the n citizen groups from the more promising policy alternatives. [Let $u_i(z)$ be the expected utility to group i from Alternative z .]

In the case of S , we are trying to maximize the total utility arising from a policy action. To do this, the utility gains/losses to one group of citizens must be compared with those to another.

Step 3 S--The utility functions u_i are multiplied by comparison scales, r_i , to get $u'_i = r_i u_i$. The r_i are chosen on the basis of the "relative needs" of the different constituency groups, and their purpose is to rescale the u_i so that a 1-unit change in utility to one group has the same social value as a 1-unit change to any other. [See subsection III.E.1 for a thought experiment by which the r_i can be impartially determined.]

With the comparison scales determined, the expression

$$\sum_{i=1}^n u'_i(z) \quad (3)$$

is a measure for comparing policy alternatives in terms of their total utility.

Step 4 S--Policy alternatives are ranked in terms of the value they are assigned by (3).

In the case of P , we are in effect simulating a "fair" bargaining session among the citizens. By "fair," we mean two things. First, for the purposes of the simulation, all citizens are "endowed" with the same rational bargaining skills and are unimpeded by transaction costs. Second, no individual or coalitional threats (i.e., pressure group tactics) are allowed, and in the absence of an agreement it is assumed that a policy 0 will be adopted that is predetermined to be "fair." (0 may be the status quo, "no government action," or an outcome with dire consequences for all.)

A remarkable theory of bargaining, discussed below, leads to the conclusion that citizens will agree on the alternative z^* that maximizes

$$\prod_{i=1}^n [u_i(z) - u_i(0)] \quad . \quad (4)$$

Hence, in carrying out P, the following steps are taken:

Step 3 P--A "fair" "no-agreement" outcome 0 is determined.

Step 4 P--Alternatives are then ranked in terms of the value they are assigned by (4).

III A JUSTIFICATION FOR THE USE OF EITHER S OR P IN TERMS OF FOUR FUNDAMENTAL NORMATIVE CRITERIA

III.A. Introduction

In Section III we will introduce four additional criteria that a reasonable preference aggregation procedure should satisfy. More specifically, we will show that both P and S satisfy different versions of one and the same set of four fundamental criteria.

III.B. Four Fundamental Criteria

The following four criteria offer perspectives from which we can compare and critique the two candidate aggregation functions S and P:

"Needs-respecting" social decisionmaking--When moral theorists or real-world policy analysts turn their attention to ascertaining which particular policy (e.g., waste management policy) is "fair," they inevitably end up justifying their choice in terms of the concept of the different needs of the persons affected. Of course, this is not to assert that either academicians or policy analysts do so explicitly, or in an ambiguous manner. Indeed, the concept of "relative needs" is notoriously elastic and problematic.

Impartial social decisionmaking--It is all but axiomatic that constituents' "needs" should be taken into account in social decision-making should be "impartial." In its simplest form, impartial means "unbiased." But what exactly does unbiased mean? This term is clarified later in the report; consequently, we will not discuss it here.

Individually rational social decisionmaking--To motivate this normative concept, a useful analogy can be made with the economic marketplace. As was informally conjectured by Adam Smith, and has been formally proven by modern economic theorists, in perfectly competitive markets, decentralized, individually rational behavior by

producers and consumers leads to a socially efficient (allocation) decision (i.e., the outcome of market trading is Pareto efficient). The root idea here is simply that, other things being equal, it is appealing if a criterion of "social optimality" dovetails in a congenial way with the concept of individually rational behavior. Both S and P have this appealing property.

Collectively rational decisionmaking--The idea that a preference aggregation rule should satisfy the criterion of "collective rationality" is a rather sophisticated one. It originated (at least in a formal manner) with Arrow's treatise (1963) on social choice theory. The problem with this normative desideratum lies in the ambiguity of the very concept of "collective rationality." Two formulations of this concept are shown in section III.F below; therefore, this matter is not discussed here.

This concludes our introductory discussion of the four basic normative criteria that a "reasonable" preference aggregation procedure might satisfy. We now turn to the principal task at hand in our research, namely the job of explicating these four important concepts, of rendering them unambiguous; and of showing how both S and P embody all four of these criteria, albeit in different guises.

III.C. "Needs-Respecting" Social Decisionmaking

The appeal to citizens' "needs" is as ubiquitous as it is ambiguous in public policy analysis. Rather than launching forth on an exhaustive discussion of the myriad senses in which the concept of needs has been used in normative theory, we proceed in a more direct and expeditious manner. First, in section III.C.1 we introduce a very precise and unambiguous characterization of the concept of "relative needs." We then show that only the procedure P satisfies this characterization. Then, in III.C.2 we introduce another, somewhat less precise concept of needs-respecting preference aggregation. We then show that S satisfies this characterization of relative needs.

III.C.1. Allocation of Utility in Direct Proportion to Relative Needs: P Aggregation

Consider a simple two-person social decision problem. For simplicity, let us assume that the decision problem concerns how two people can "fairly" divide a pie. The prospect space, consisting of all feasible utility n -vectors (with $n = 2$ now) for such a problem, is sketched in Figure 1. It is assumed that the two utility functions have already been interpersonally calibrated with respect to both the origin and the unit of the utility functions.* Later, it will become clear in what sense--and why--this assumption of interpersonal comparisons can be discarded, at least at an operational level. The origin of the prospect space u will be denoted by the vector d , and is assumed to correspond to the "no-agreement" outcome. Note that the Pareto boundary in Figure 1 is flat. We have intentionally started off with this special case.† When there is a flat boundary, there is an unambiguous measure of relative needs. This is the slope of the Pareto frontier. In the present example, player j gets 2 units of interpersonally calibrated utility for every unit that i gets. That is, there is a 2:1 needs ratio. Of course, in asserting this, we are in effect equating "relative needs" with "relative intensity of desire." This seems unobjectionable because our concern is with relative--not absolute--needs.

FIGURE 1

*In section III.E we sketch a method by means of which the units of n utility functions can be calibrated.

†In the context of a pie division problem where there is a continuously divisible commodity, this case will correspond to an assumption that both players have constant marginal utility for pie.

In order to determine an allocation of pie that corresponds to the relative needs of i and j , let us introduce our fundamental norm of equitable distribution:

Postulate A: Distribution in Proportion to Relative Needs--Pie should be distributed such that the utility gains to the recipients are proportional to their relative needs. Formally,

$$a_i(u_i - d_i) = a_j(u_j - d_j) \quad , \quad (5)$$

where the variables a_i , a_j correspond to the slope of the Pareto frontier. In the present case, we require that $a_i:a_j = 2:1$.

Postulate A in and of itself is not sufficient to determine a unique solution to the problem, because if the reader consults Figure 2, he will observe that any and every point on the line CD satisfies (5). Therefore, we supplement our Postulate A by a straightforward:

Postulate B: Efficiency--The pie must be distributed in such a way that an efficient (Pareto optimal) outcome is achieved. Formally, we require

$$\text{MAX}_{U_k} \sum_k a_k u_k \quad k = i, j \quad . \quad (6)$$

Note that this is not the ordinary efficiency condition. For the a_k here are not variables, but are the given slope values, a_i , a_j . Choice of any other values for these weights would clearly render (6)

The equation of CD is (5) above;
The equation of AB is $[a_k u_k = c$
(where c is a hyperplane
constant)

FIGURE 2

incompatible with (5). Taken together, (5) and (6) clearly define a unique solution to the distribution problem, namely a utility vector u^* lying midway along the Pareto frontier. At this point, i receives a utility gain one-half as great as j receives, which seems reasonable because i 's need is half as great as j 's. Geometrically, u^* is the point of intersection of the two lines $\underline{AB}$ and $\underline{CD}$, which have the property that their slopes are equal in magnitude but opposite in sign.

This point u^* is none other than the solution given by the aggregation rule P. Specifically, we can state:

Theorem I: The vector u^* that maximizes the arithmetic product of the utility gains, i.e., the vector implied by the P aggregation procedure, coincides with the vector u^* that is the point of intersection of the lines $\underline{AB}$ and $\underline{CD}$ in the foregoing analysis.

Proof: Let $y = f(x)$ denote the equation of the Pareto frontier, where we can identify y with u_1 and x with u_2 . $f(x)$ is of course a convex function, because the boundary of the Pareto frontier will always be convex (given the assumption that jointly randomized strategies can be utilized). We need simply establish that the product $A = x \cdot y$ is maximized when $dy/dx = (y/x)$, an equation that is isomorphic to our relation (5) above.

Assume that $A_0 = x_0 \cdot y_0$ is maximal. Let us now consider the nature of areas A that are neighboring to the maximal area A_0 . Specifically, consider Δx , where Δx is assumed small. Clearly

$$\Delta A = (x_0 + \Delta x) \cdot (y_0 + \Delta y) - x_0 y_0 = 0,$$

given our assumption of the maximality of A_0 . But $\Delta A = y_0 \Delta x + x_0 \Delta y$ (assuming we ignore second-order terms). Hence,

$$-y_0 \Delta x = -x_0 \Delta y \rightarrow (\Delta y / \Delta x) = -(y_0 / x_0).$$

Passing to the limit gives the desired result. ∇

Three remarks are in order about the way in which P aggregation respects the norm of allocation in direct proportion to the parties' relative needs. First, it should be noted that Postulate A above stipulates that utility--not "pie"--is the fundamental object of distribution. Pie (e.g., health effects in our NRC example) is only important because it is a source of (i.e., it generates) utility. To state the matter somewhat differently, pie is important because it fulfills humans' needs and therefore is a source of social welfare. Hence, to the extent that

something must be apportioned in accordance with a mathematical formula such as relation (5) above, it is utility that is so allocated in the theory we have discussed.

Second, note equation (6)--the Pareto efficiency condition peculiar to P aggregation. The weights a_1, a_j here are not constant, but rather depend upon the particular example being analyzed. As we have seen, these weights are the coefficients of the players' relative needs (relative intensity of desire); and clearly needs differ from one case to another case. We can thus see that use of P aggregation is equivalent to maximization of a needs-weighted sum of the parties' utilities. This is an interesting and significant observation. [Note the contrast to the situation in S aggregation where (by definition) one maximizes the un-weighted sum of the already calibrated utilities.]

Third, we should say something about the use of interpersonal comparisons of utilities in P aggregation. Above, we assumed that the unit and origin of the utility scales were interpersonally calibrated. It was necessary to do so for conceptual purposes. Our purpose was to discuss the way in which a certain social decision model (namely P aggregation) incorporates the concept of allocation in proportion to relative needs; and clearly any discussion of relative needs implies the use of interpersonal comparisons, at least at a conceptual level. However, Theorem I above places things in a somewhat new light. It should be recalled from our discussion above that P aggregation (unlike S aggregation) has the very special property of being invariant under separate linear transformations of the players' utilities; that is, the physical solution to the social decision problem (e.g., the optimal criterion for acceptable risks) will not depend upon operational interpersonal comparisons of utility.*

But how can this fact be squared with what we have observed above, namely that if we mathematically restate the solution generated by P aggregation in the form of equations (5) and (6), we see that with

*The distinction between "conceptual" and "operational" interpersonal comparisons in the context of P aggregation is discussed at length in Brock (1979), where a novel derivation of P aggregation is given.

P aggregation, utility is distributed in proportion to relative need? The answer lies in one simple but important observation. Because of the invariance of P aggregation noted in Section II, and because of Theorem I, we know that equations (5) and (6) will hold for all possible rescalings (calibrations) of the players' utilities. But this means that in particular, (5) and (6) will hold for that unique calibration that corresponds to the "true" interpersonally calibrated utility scales--scales so difficult to "find" in practice. We can summarize all this by stating that P aggregation does not require the making of interpersonal comparisons of the parties' utility scales at an operational level. However, interpersonal comparisons may be said to arise at a conceptual level, only if we hypothesize the existence of a very special calibration that makes interpersonal judgments meaningful.

An ostensible limitation of the foregoing account is that it assumed a "flat" Pareto frontier so that a well-defined measure of relative needs exists. In general, the Pareto frontier will not be flat, rather it will be (piecewise) convex throughout. Fortunately, the proposed theory extends to this more general case. Let us see how this is so.

The reader can see by inspection that when the Pareto boundary of the payoff space is strictly convex (see Figure 3), there is once again a unique solution to our relations (5) and (6). That is, there is a unique point of intersection of AB and CD. However, two salient differences exist between this more general case and the "flat" case considered earlier. First, it is no longer meaningful to suppose that the

FIGURE 3

efficiency condition (6) merely "supplements" the fair-sharing condition (5). Rather, both conditions enter on equal footing, and the values of the variables a_i , a_j , as well as of the payoffs $(u_i - d_i)$, $(u_j - d_j)$ are determined simultaneously when (5) and (6) are solved. This is to be contrasted to the earlier case where the values of a_i and a_j were known from the outset. A second difference is that whereas we originally had a global measure of relative need, we now have only a local measure, namely the slope of the line AB tangent to the solution u^* . However, we can justify an identification of this local measure with a global measure on the following grounds. For any and every strictly (or piecewise) convex problem U there will be a unique flat problem $\bar{U}$ constructed as follows. Simply replace the original problem U with the problem $\bar{U}$ whose prospect space is the set of utility vectors bounded by the line AB, which is tangent to the solution u^* to the original problem. This flat problem $\bar{U}$ will clearly have the same solution u^* as the original convex problem, so that the two problems can be deemed solution equivalent. And the measure of relative needs in both problems will clearly be the same, namely the slope of the line AB. We can therefore let this slope serve as a global measure of relative need in both U and $\bar{U}$.

Note the important role of the disagreement payoff d in the proposed theory. The solution a^* and u^* to (5) and (6) clearly will depend upon the parameter d . To point this out does not prejudge whether there exists a meaningful zero point in a given problem. Often there will not be one. However, this does suggest that if we want an ethical theory that is firmly based upon an unambiguous account of relative needs, then a zero point will necessarily enter into the matter at some point.

III.C.2. Allocation in Approximate Accord with Relative Needs: S Aggregation

When we turn to S aggregation, the sense in which social divisions can be said to be needs-respecting is somewhat less precise than it is in the case of P aggregation. We can best clarify the matter by distinguishing two different classes of empirical situations:

Case A: The Case of Constant Marginal Utility (linear utility functions)--In the event that all citizens have linear utility functions (e.g., constant marginal utility), the utility payoff space of a social decision problem will be bounded by a flat (straight line) Pareto boundary, as illustrated in Figure 4 below. In the particular situation depicted in Figure 4, party i has an ever-so-slightly greater need for what is at stake than party j because if he prevails and gets his way completely, then he gains $10 + \epsilon$ utils. The S aggregation rule will in this situation (and in all flat cases) award all the utility at stake to party i, the needier. That is, i will get all the pie and j will get none. In this sense, the S rule can be said to be needs-respecting. [The P-rule will of course select a compromise outcome; namely, the utility vector lying midway along the Pareto frontier, an outcome which awards slightly more utility to i (who is slightly needier)].

FIGURE 4

Case B: The General Case--In general, the parties will not have linear utility functions, and the boundary of the payoff space will be strictly convex, as in Figure 5. In this case, S aggregation selects as the solution the utility vector u^* with the following property: the hyperplane tangent to the Pareto boundary at the solution point will have a slope of -45° . (This follows from the fact that the iso-social utility contour of the S rule is given by the equation $\sum u_i = c_i$ where c is a hyperplane constant.) What this means is that the solution is the point at which the parties are equally needy, in a local sense.

FIGURE 5

That is, if $W \triangleq \sum u_i$, then $dW/du_i = dW/du_j$ at the solution. This is ostensibly all that can be said about the sense in which the S aggregation rule is needs-respecting in the general sense.

Because of the rather different way in which S aggregation is needs-respecting in Case A situations and Case B situations, it cannot be asserted that S aggregation is needs-respecting in as unambiguous a way as is P aggregation. Nonetheless, S aggregation certainly pays some heed to the concept of need--at least to certain characterizations of that all-too-elastic concept!

III.D. Impartial Social Decisionmaking

The concept of "impartiality" is so widespread and popular that it is often all but identified with "fairness." A problem that confronts us at the outset, however, is that of making clear exactly what "impartial" means. In this section of the report, we shall attempt to show that both S aggregation and P aggregation respect the intuitively compelling concept of impartiality. However, each of these two aggregation procedures embodies decidedly different versions of the notion of impartiality. In section III.D.1, we consider the sense in which P aggregation is impartiality-respecting; in section III.D.2 we look at S aggregation.

III.D.1. P Aggregation and the Concept of Impartial Social Decisionmaking

First, let us introduce one particular characterization of the concept of impartiality in social decisionmaking. We need a definition:

Definition: Consider the role of a planner or policymaker who is charged with making a decision that will have implications for two or more other people (e.g., citizens). We shall say that he is acting impartially if the only information that he takes into account in arriving at his decision is ethically significant information. More specifically, the only information he takes into account is information concerning the differential needs of the parties for what is at stake.

The idea here is quite a simple one. We would hardly call the decision of a planner impartial if it reflected information about the telephone numbers (or waistline measurements) or physical appeal of the citizens whose interests he is arbitrating. Rather we would call such a decision biased. However, if the only information he takes into account is information bearing on differences in the citizens' likes, dislikes, and needs for what is at stake, we would be inclined to say that his decision was an impartial one. (Of course, this presupposes that he makes good use of this ethically significant information--i.e., that he makes use of a compelling aggregation rule. It is conceivable that a planner could possess ethically pertinent information, but use it in an absurd way.)

With this point in mind, let us now consider P aggregation. Our foregoing analysis showed that use of the P aggregation rule is tantamount to using the social welfare function

$$\text{MAX}_U \sum a_i u_i ,$$

where the weights a_i are not arbitrary, but rather are the coefficients of the relative need of the parties for what is at stake in the decision problem. [Recall equations (5) and (6) above.] But this being the case, it follows from our above definition of impartiality that P aggregation is impartial, because it is tantamount to aggregating the needs-weighted utilities of the members of society. To the extent that the interests of party i receive a different weight from the interests of party j (i.e., to the extent that $a_i \neq a_j$), there is a good reason for the

asymmetry: the needs of the two parties are different. It is perhaps worth pointing out that most decisions taken in daily life are not impartial in the sense we have just outlined. Such decisions are more often than not the result of multilateral bargaining between coalitions and interest groups--whether within a family, a church, or a body politic at large. Now it can be shown in the mathematical theory of games (Harsanyi, 1963, 1977; Brock, 1978, 1978A, 1978B, 1979) that the outcome of bargaining will in general satisfy an equation of the form

$$\text{MAX } \sum b_i u_i \quad (7)$$

where the coefficients b_i summarize three sources of strategic differences among the parties; namely, differences in relative need for what is at stake; differences in unilateral threat power; and differences in coalitional threat power. As we have argued elsewhere (Brock, 1978), only the asymmetries implicit in category (i) are ethically respectable and should be reflected in a social welfare function as we ordinarily interpret that concept. In contrast to this, (7) can best be called a social power function (Aumann and Kurz, 1977).

III.D.2. S Aggregation and the Concept of Impartial Social Decisionmaking

S aggregation will now be shown to embody a somewhat different, and in fact significantly stronger concept of impartiality than that of P aggregation. This fact was first observed by the game theorist John C. Harsanyi who has given a very deep and interesting justification for S aggregation (Harsanyi, 1953, 1955). Just as we did above, let us start off with a definition.

Definition: Let us say that an aggregation rule is impartial if it reflects the preferences of a person who is impartially situated in the following very specific sense: he does not know his identity in the decision problem at hand. Rather, he simply knows that he has an equal chance of being in the shoes of any of the people affected by the decision that is adopted. To state the matter somewhat differently, he is impersonally situated.

Harsanyi has shown* that the utility function (or "individualistic social welfare function") of a person who is impersonally situated will be none other than the familiar Utilitarian form

$$\text{MAX } \sum u_i \quad , \quad (8)$$

which the reader will at once recognize as our S aggregation rule. Let us be clear on what is being asserted here. Equation (8) will serve as the "utility function" of a given person who is impersonally situated, in the sense that he will evaluate the alternative prospects accessible to him (and to any of the other people whose role he may assume) by means of (8). Why is this so? Because as Harsanyi shows quite clearly, it is rational for an individual who is impersonally situated to use (8) to evaluate the outcomes of alternative policies that affect the citizens of the society of which he is a member. In short, Harsanyi identifies impartiality with impersonal rational decision. As we have seen by our discussion in section III.D.1, there is at least one alternative way of characterizing the concept of impartiality. The reader can best judge for himself which he finds more persuasive. In the last analysis, this matter may hinge upon the particular application intended for an aggregation procedure.

III.E. Individual Rationality of S and P

We have just completed a discussion of impartial social decision-making. We now analyze the way S and P aggregation can be integrated from the standpoint of individual rationality.

III.E.1 Individual Rationality of S

In what circumstances would an individual adopt S as his decision rule? Imagine a fictitious psychologist (FP) who (a) has to decide on a policy $z \in Z$, (b) strives to be rational in the sense of VNM, and (c) has an equal chance of becoming any one of the citizens after making

*Kenneth Arrow has pointed out that Harsanyi's argument here (Harsanyi, 1953) was independently formulated at an earlier date by William Vickrey.

his decision. (The uncertainty about his future identity forces our imaginary psychologist to be impartial.) Suppose further that FP is given a description d_j of each citizen j , as well as a VNM representation u_j of j 's preferences P_j .

How will FP decide which policy alternative to pick? The "extended outcomes" he faces are of the variety $(d_j; z)$; i.e., he becomes person j and experiences policy effects z . Let v represent FP's VNM preferences over these extended outcomes. If FP knew v , then determination of an optional policy would be straightforward. FP would (by assumption of "rationality") simply select the policy that maximized $v(z)$. The problem lies in assessing v in the first place. In arriving at v , FP is forced to draw on his knowledge of human beings in different conditions and to think in terms of psychological laws and "universal human values." Let us now formally investigate the relationship between v and the VNM preferences $\{u_j\}$ of the citizens.

Consider the function that is defined from v by holding the person j constant:

$$u'_j(z) = v[(d_j; z)] \quad . \quad (9)$$

Given that the same basic psychological laws apply to everyone, it seems reasonable to assume that the structure of FP's VNM preferences among the z , given that he is j , ought to be the same as that of j 's preferences (after careful thought and consultation). In other words, we are assuming that j 's preferences are the natural ones to have if one is j --in effect that these are determined by universal psychological laws, and that this fact must be reflected in v . Hence u'_j is simply a VNM representation of P_j , namely

$$u'_j \sim_j u_j + b_j \quad , \quad (10)$$

* Our approach here of an "impartial" FP differs from an alternative approach whereby a government policymaker's own personal preferences are elicited and used in policy analysis.

where r_j is a positive scaling factor and b_j is a shift of origin, both parameters being chosen on the basis of interpersonal comparisons of utility (see below).

Now let us return to the question of how our rational psychologist will make his decision. For each policy alternative, z , he will compute his expected utility if he becomes citizen j , i.e., $[v|z; d_j]$. He will then compute the overall expected utility of z , $[v|z]$, by taking into account that he could become any of the citizens with equal probability.

In symbols, the results are

$$[v|z] = \left(\frac{1}{n}\right) \sum_{j=1}^n [v|z; d_j] \triangleq v(z) \quad (11)$$

However, because of (9) and (10) above,

$$[v|z; d_j] = [u'_j|z] \triangleq u'_j(z) \quad (12)$$

Thus,

$$v(z) = \left(\frac{1}{n}\right) \sum_{j=1}^n u'_j(z) = \frac{1}{n} \sum_{j=1}^n r_j u_j(z) + \sum_{j=1}^n b_j \quad (13)$$

Being VNM-rational, FP will pick the policy, z^* , which maximizes $v(z)$. (This is equivalent, however, to picking the z^* that maximizes $v(z) - \sum b_j$.) Thus, FP's decision rule has the same additive form as S.

What remains to be shown is that the r_j coefficients in the above formalism are the interpersonal "calibration factors" reflecting the different citizens' differing "needs."

III.E.1.a. Interpersonal Comparison Coefficients

We propose to give a simple example that demonstrates how the r_j in FP's utility function (13) correspond to our intuitive understanding of "relative need," and hence serve as the interpersonal comparison

coefficients (or scale factors) that are required by S aggregation* (recall II.D above).

For simplicity, suppose that there are only two citizens and two policy effects. The first citizen is a typical youth from India. The second is an offspring of a well-known billionaire. The two potential policy effects are: a citizen receives \$1 and a citizen receives \$50,000.

Thus, the extended alternatives are:

- (I) Indian youth receives \$1.
- (II) Indian youth receives \$50,000.
- (III) Rich youth receives \$1.
- (IV) Rich youth receives \$50,000.

Let the Indian's VNM preferences be represented by u_1 and the rich youth's by u_2 . We assume that both prefer receiving more money to less money and that the u_1 have been scaled so that

$$u_1(a) = u_2(a) = 0; \quad \text{and} \quad u_1(b) = u_2(b) = 1 \quad (14)$$

Now imagine that FP has to come up with a v representing his VNM preferences over I to IV. We assume that FP prefers IV to III to II to I. Because V is unique only up to a positive affine transformation, we can scale v so that $v(\text{IV}) = 1$ and $v(\text{I}) = 0$.

To determine $v(\text{II})$ and $v(\text{III})$, FP compares II and III with gambles between I and IV. For what value of the probability, p , would FP be indifferent between having II for sure, and having to face the gamble--IV with Probability p and I with probability $1-p$?

*We intentionally use the notation r_j instead of a_j here because the a_j variables introduced in III.C.1 above will not generally coincide with the r_j variables. This matter is discussed at length in Brock (19788).

A reasonable answer might be $p = 0.96$. Then according to the VNM theory, $v(\text{II}) = 0.96$. Now, what about III for certain versus the above gamble?

For this case, a typical value might be $p = 0.998$. Thus $v(\text{III}) \approx 0.998$.

Finally, suppose there are four policy alternatives, each of which will produce one of the possible combinations of policy effects:

	<u>Indian (dollars)</u>	<u>Rich Youth (dollars)</u>
z_1	1	1
z_2	1	50,000
z_3	50,000	1
z_4	50,000	50,000

Because FP has an equal chance of being either youth, he would evaluate the expected utility of the z_i as follows:

$$\begin{aligned}
 v(z_1) &= 1/2 v(\text{I}) + 1/2 v(\text{III}) = 0.499 \\
 v(z_2) &= 1/2 v(\text{I}) + 1/2 v(\text{IV}) = 0.5 \\
 v(z_3) &= 1/2 v(\text{II}) + 1/2 v(\text{III}) = 0.979 \\
 v(z_4) &= 1/2 v(\text{II}) + 1/2 v(\text{IV}) = 0.98
 \end{aligned}
 \tag{15}$$

By the result (13) above, v can be written as some linear combination of u_1 and u_2 .

$$v(z_i) = [r_1 u_1(z_i) + b_1] + [r_2 u_2(z_i) + b_2] \tag{16}$$

Equating (15) and (16) and solving for the r_j and b_j , we get

$$\begin{aligned} r_1 \text{ (Indian)} &= 0.480 \\ r_2 \text{ (Rich)} &= 0.001 \\ b_1 + b_2 &= 0.499 \end{aligned} \tag{17}$$

(The b_j cannot be determined exactly because of redundancy in the equations.)

Thus FP would weigh the Indian's utility by a factor about 500 times larger than that for the rich youth--a result that we believe corresponds to and makes precise an intuitive judgment about the youths' "relative need."

The preceding example involved an extreme discrepancy in wealth. In practice, citizen groups will not be so different from each other as the Indian youth and the offspring of a billionaire. Hence, the interpersonal comparison coefficients would turn out to be much closer in value. In situations where there is no reason to believe one group is better off than another, the coefficients would all be equal.

III.E.2. Individual Rationality of P

As pointed out above, P is equivalent to frictionless bargaining among rational individuals. The most elegant and well-known derivation of this result is due to Nash (1950); however, it is somewhat abstract. We will sketch a more heuristic derivation, due to Zeuthen and Harsanyi, that is based on a concrete model of the bargaining process (see Harsanyi, 1977, chap. 8).

III.E.2.a. Bargaining Model

For simplicity, we consider the case where there are only two players: players 1 and 2. We assume the bargaining takes place in stages and is governed by the following rules.

At each stage k , either player has three alternatives:

- b1: He may repeat the offer he made at the previous stage a_i^{k-1}
- b2: He may accept in full his opponent's last offer, a_j^{k-1}
- b3: He may make a new offer, a_i^k , that is more favorable to his opponent than a_i^{k-1} , but less favorable to himself (and of course more favorable to himself and less favorable to his opponent than his opponent's last offer, a_j^{k-1}).

The bargaining terminates when either player accepts his opponent's last offer, b2, or when both players stand on their last offer, b1, b1. In the first case, the players get what they have accepted; in the second, the bargaining collapses and both players get the best they can achieve without an agreement--a state we will call 0.

III.E.2.b. Zeuthen's Principle

To complete the model we need a theory of concession making. Intuitively, what causes a player to make a concession is a fear that he might end up with 0 when his opponent might agree to something better.

Zeuthen formalized this intuition into the following principle: the player who will make the next concession is the one who can least tolerate the chance of there being no agreement.

In calculating such a tolerance level, we simplify the bargaining situation so as to more easily arrive at a measure of each player's willingness to stick to his terms. Suppose the only alternatives open to each player are b1 (stand on last offer and chance a conflict) or b2 (accept opponent's last offer). Let p_{21} be Player 1's subjective probability that Player 2 will stick to his guns if 1 does also. For 1 rationally to stand on his last offer, his expected utility must be at least as great as if he concedes, i.e.,

$$(1 - p_{21}) u_1(a_1) + p_{21} u_1(0) \geq u_1(a_2) \quad , \quad (18)$$

where a_i is Player i 's last offer ($i = 1, 2$).

But this implies that

$$p_{21} \leq \frac{u_1(a_1) - u_1(a_2)}{u_1(a_1) - u_1(0)} \triangleq r_1 \quad (19)$$

Note that r_1 can be interpreted as the largest chance that Player 1 can tolerate that Player 2 will decide on b_1 . In other words, r_1 is the largest chance 1 can take of there being no agreement, and hence, r_1 is sometimes called Player 1's "risk limit." The r_1 can also be interpreted as the ratio of the cost of giving in to an opponent's terms to the cost of no agreement. Thus, it measures Player 1's incentive to stick to his last offer.

Zeuthen's principle is simply that the player who is less willing to face a conflict (i.e., who has the lesser r_i) will make the next concession.

To see this more vividly, image two hypothetical bargainers staring each other in the eye, both believing at first that the other is likely to give in (Low p_{ij}). Gradually, as neither does, their estimates of p_{ij} will increase until the player with the lower r_i decides $p_{ij} \geq r_i$ and so gives in. (We assume here that each player has equal bargaining skill and resources.) In general, bargainers have an intuitive sense of the r_i in terms of "who would lose the most if there's no deal." This puts psychological pressure to give in on the one with the lowest r_i .

III.E.2.c. Derivation of P from Zeuthen's Principle

Observe that

$$r_i \triangleq \frac{u_i(a_i) - u_i(a_j)}{u_i(a_i) - u_i(0)} \leq \frac{u_j(a_j) - u_j(a_i)}{u_j(a_j) - u_j(0)} \triangleq r_j \quad (20)$$

is equivalent to

$$\begin{aligned} P(a_i) &\triangleq [u_i(a_i) - u_i(0)] \cdot [u_j(a_i) - u_j(0)] \\ &\leq [u_i(a_j) - u_i(0)] \cdot [u_j(a_j) - u_j(0)] \triangleq P(a_j) \end{aligned} \quad (21)$$

Thus Player 1 will make concessions, reducing his utility gain--
 $[u_1(a_1) - u_1(0)]$ --and increasing his opponent's $[u_j(a_1) - u_j(0)]$ until
 $P(a_1) \geq P(a_j)$. At this point, j has to start making concessions and so
 on. Gradually $P(a)$ increases, and if we make the realistic assumption
 that each concession must raise $P(a)$ by some minimum amount, then even-
 tually one of the players must make the offer a^* which maximizes P over
 the set of possible agreements.

Now if either player were to make a concession beyond a^* , then he
 could always renege and press his opponent back to a^* . Hence, it would
 be pointless for either player to make such concessions. Once both
 players ascertain the stability of a^* , it behooves them to agree to it
 because this is the best each can do.

In sum, the outcome of rational, frictionless bargaining is the same
 as that selected by the social decision rule, P . This result is the same
 no matter whether we model the bargaining as Zeuthen did or whether we
 follow Nash in deducing the outcome from abstract a priori postulates.

III.F. Collective Rationality of S and P

We now shift our attention to the concept of "collective rationality"
 and interpret S and P in this new light.

III.F.1 Collective Rationality of S

We have shown that S would be adopted by a hypothetically rational
 individual who is impartially situated. We next ask whether there are
 principles of collective rationality that would lead citizens to adopt S.

The answer is yes. The postulates are the following:

- (I) Citizen Rationality--Each citizen's preferences are rational
 in the sense of VNM.
- (II) Dependence on Interpersonal Cardinal Utility--The procedure
 is to be based on interpersonally-calibrated VNM utility
 functions.
- (III) Collective Preference Rationality--Social preferences are to
 be rational in the sense of VNM.

- (IV) Weak Responsiveness of Social Preferences to Citizen Preferences--If all citizens are indifferent between A and B, then society should be indifferent between them as well.
- (V) Strong Symmetry--The procedure should be symmetric in its utility arguments--i.e., it should assign equal weight to each u_i .

Postulates (I), (III), and (IV) imply that the social decision rule must have the same additive form as S. For a proof see Harsanyi (1955). Postulates (II) and (V) imply that the weights in the sum should all be equal to unity (Harsanyi, 1977), once the utility functions have been interpersonally calibrated as in III.E.1.

Postulates (I), (IV), and (V) seem unobjectionable. Postulate (II), of course, is a stronger form of citizen sovereignty, but both (II) and (V) embody the ethical judgment that social decisions should be based on need. Postulate (III), however, raises some questions. Why should the principles of individual rational decisionmaking be adopted in the collective context also?

Let us examine the VNM rationality principle anew in a normative social context and see whether they are still reasonable. In particular, let us consider Marschak's formulation of the rationality principles in terms of his postulates PI, PII, PIII', and PIV (Marschak, 1950).

PI requires that the social ordering of policy alternatives be complete and transitive. We need completeness in order to come to a final decision. As for transitivity, if society preferred A to B, B to C, but C to A, an outsider could charge the society for changing A to B, and then B to C, and then C back to A, thereby leaving society where it started, except a little bit poorer. Thus, in order to rule out this kind of "dynamic" irrationality, it is reasonable to require that the social ordering be transitive. Hence, PI appears to be as compelling a decision requirement for society as for the individual.

But the same applies for the other principles also. If society prefers A to B to C, a requirement that society be indifferent between B and some gamble between A and C does not seem at all unreasonable. Hence, PII can still be adopted. PIII' says simply that there are at least four mutually nonindifferent prospects. Lastly, PIV stipulates that if society is indifferent between A and A', then it should also be indifferent between gambles in which A is substituted for A'. This seems reasonable.

The Marschak postulates are often objected to as a theory of individual rationality on the grounds that they rule out the "love of danger" or "sin" associated with taking chances (see Sen, 1970, pp. 96-7). We submit, however, that although these objections may be valid if we are trying to describe individual behavior, they carry no weight in the normative social context. In a separate vein, Kenneth Arrow (1963, p. 10) has claimed that the use of VNM utility is inappropriate because it makes social decisionmaking dependent upon "the tastes of individuals for gambling." This objection, however, misses two fundamental points. First, although an individual's VNM utility function is usually encoded by giving him a choice between gambles, it contains important information about the strength of his preferences--information that cannot be left out of equitable social decisionmaking. Second, even the government cannot eliminate uncertainty, and so citizens' "tastes for gambling" should be incorporated into the social decision procedure. Basically, Arrow seems to have misunderstood the importance of using VNM utility functions in satisfying the *desiderata* a and b introduced at the beginning of this paper.

Before turning to P, we note that postulates similar to (I), (III), and (IV) imply that the individual moralist can represent his ethical preferences as some weighted sum of citizens' individual utility functions. The modification required is simply one of perspective. Postulate (I) is unchanged. Postulate (III) becomes the condition that the moralist's ethical preferences satisfy the VNM axioms, and Postulate (IV) that his ethical preferences are indifferent between two alternatives whenever all the citizens' are. For a proof of the additive result for the moralist see Harsanyi (1977).

III.F.2. Collective Rationality of P

Finally, we consider whether there are any appealing collective rationality criteria that lead to P as a social decision rule.* Along these lines, Kaneko and Nakamura (1979) have proposed the following postulates. In addition, they have shown that these postulates are necessary and sufficient conditions for alternatives to be ranked socially according to the P aggregation rule.

- (I) Existence of Origin--There exists some alternative that represents one of the worst states for all citizens (e.g., everyone dies, or no more steps are taken to safely store high-level nuclear wastes). This alternative is designated the "origin," and social welfare is evaluated by considering the relative increases of individuals' welfare from this position.
- (II) Pareto Optimality--If all citizens hold that Alternative A is at least as good as Alternative B, and at least one citizen strictly prefers A to B, then A should be strictly preferred to B in the social ranking.

* The reader satisfied with the arguments for S may wonder why we raise the question at all. How can there be any satisfying criteria leading to P? They must contradict those leading to S. In answer to such a reader, we point out an objection to S due to Diamond (1967).

The objection is that S is sometimes indifferent between alternatives when there appears to be a clear reason for one being socially preferable to another. Consider Diamond's example.

Suppose there are two citizens C1 and C2, two policy alternatives A and B, and two possible random states of nature, θ_1 and θ_2 . If A is chosen, let C1 get utility 1 and C2 get 0, no matter whether θ_1 or θ_2 occurs. If B is chosen, again let C1 get utility 1 and C2 get 0 if θ_1 occurs; but if θ_2 happens, let the utility assignments be just the reverse. In table form

	<u>If θ_1 occurs</u>	<u>If θ_2 occurs</u>
Alternative A	$u_{c1} = 1, u_{c2} = 0$	$u_{c1} = 1, u_{c2} = 0$
Alternative B	$u_{c1} = 1, u_{c2} = 0$	$u_{c1} = 0, u_{c2} = 1$

- (III) Independence of Irrelevant Alternatives (with the Neutral Property)--The social preference between two alternatives depends only on the individuals' preferences between these alternatives, regardless of individuals' preferences concerning other alternatives; and social preferences do not depend on the labeling of alternatives.
- (IV) Anonymity--Social preferences do not depend on the labeling (naming) of individuals. (This condition is identical to Desiderata c above.)
- (V) Continuity--If A is strictly preferred to B under the social preferences, and B to C, then there exists a gamble between A and C that is socially indifferent to B.

The reader is referred to Kaneko and Nakamura's original paper for the rather technical proof of this result.

The postulates are self-explanatory, except perhaps for (I). Kaneko and Nakamura's approach is equivalent to simulating bargaining, except that the "no-agreement" bargaining outcome is designated as one of the worst conceivable states for everybody. The effect of postulating such

(We have assumed that u_{c1} and u_{c2} have been scaled so that their coefficients in the sum, S , are 1.) Clearly, $S(A) + S(B) = 1$. However, Diamond objects that B seems strictly preferable to him "since it gives C2 a fair shake while A does not."

A possible response to Diamond is to add an additional axiom to those leading to S . The purpose of the axiom would be to "break ties," like those in the above example, in favor of the alternative that most evenly distributes expected utility (e.g., choose the alternative that minimizes

$$\sum_{i=1}^n \left| 1 - \frac{s^*}{n \cdot u_i} \right| ,$$

where s^* is the value of the alternatives with maximum total utility). Another response is to look for a new set of rationality postulates and another decision rule (e.g., P) that distributes utility gains more equally among the citizenry. This is what Kaneko and Nakamura have done.

a dire outcome is to make everyone equally interested in making concessions, including those who are well off. Kaneko and Nakamura propose that the bargaining is more equitable if citizens are "leveled" in this way.

III.G. Conclusion

In Sections I-III of this report, we have surveyed two methods of preference aggregation: S (referring to summation), and P (referring to multiplication). We have endeavored to show how both S and P are consistent with four fundamental criteria of satisfactory social decision-making:

- "Needs-respecting" social decisionmaking
- Impartiality
- Individual rationality
- Collective ("social") rationality.

More specifically, we observed that S and P are compatible with different versions of these four abstract normative criteria.

In the remaining portions of the report, we turn to some of the practical problems that arise when we try to implement a preference aggregation procedure in practice. Specifically, we discuss how to apply S aggregation in the context of the nuclear waste management problem.

IV THE APPLICATION OF S IN THE MEASUREMENT OF THE PERFORMANCE OF NUCLEAR WASTE MANAGEMENT SYSTEMS

In the preceding sections, we discussed two methods for analyzing and ranking public policy alternatives. These were maximizing the sum and the product of citizens' utility gains, respectively. The first method, which we labelled S, requires the interpersonal comparison of utility. The second method, P, entails simulating fair bargaining among the citizenry. In carrying out P, no interpersonal comparison of utility is necessary. However, in the case where bargaining leads to "no agreement," the citizens' utility gains must be determined. In the case of nuclear waste, presumably "no agreement" means maintenance of the status quo.

The purpose of Section IV is to show how to apply S in developing a performance measure for nuclear waste systems. Several reasonable assumptions that allow a great simplification in the analysis will be introduced in this section. The development of the performance measure will be reduced to the determination of several social tradeoff parameters concerning statistical health effects, safety expenditures, and present effects versus future effects.

In the next section, we will turn our attention to estimating the tradeoff parameters. This could be done by surveying citizens' preferences concerning the health effects and safety expenditures in their own lives. However, this data is not currently available. Another approach is to infer the parameters from expert and laymen's direct judgements about the parameter values. This is the task we will carry out, applying Bayesian inference techniques to some sample judgements.

The application of P awaits information about the risks of "no agreement" in waste management policy. It is a possible topic for future work.

IV.A. The Analytical Problem Simplified

S is really a modern form of utilitarianism ("the greatest good for the greatest number"), an ethical philosophy that dates back to the work of Jeremy Bentham, James Mill, and John Stuart Mill. Its modern quantitative form owes much to the work of J. Von Neuman and O. Morgenstern (1947), which struck a balance between a purely cardinal and a purely ordinal theory of individual preferences, and to the work of J. Harsanyi (1977), which restored the objective nature of interpersonal comparisons of utility.

To apply S, the total expected utility change of each major policy alternative must be estimated and compared. To accomplish this the expected utility of the most significant outcomes of a policy must be computed. A decision tree, illustrated in Figure 6, is a graphic method for analyzing policy in order to make the necessary estimates and comparisons.

Going from left to right in Figure 6, the first branches are associated with the different waste management alternatives under consideration (e.g., methods of encapsulating high-level nuclear waste). We have labelled one of the alternatives (1). The others are schematically symbolized by the single downward branch.

The rest of the branches depict the sequence of the most important events that might occur if the i^{th} system were chosen. For instance, a certain type of accident might occur, such as the flooding of an underground storage chamber. Let j stand for a particular type of accident. The probability or likelihood that j will happen to system i we have called p_{ij} .^{*} As a result of the accident, the health of several people might be affected in various ways. Let k stand for a particular scenario of health effects in which h_k people are affected in the l^{th} manner (e.g., cancer). In addition, let q_{ijk} be the best estimate of the probability that scenario k will follow from accident j .

^{*} p_{ij} might be the best consensus estimate of a group of unbiased experts.

FIGURE 6

SCHEMATIC DECISION TREE ANALYSIS OF NUCLEAR WASTE MANAGEMENT POLICY

Ultimately, we are concerned with peoples' preferences (utility) between potential health effects on the one hand and money expenditures to reduce the risks or benefits that can be obtained only by incurring the risks, on the other hand. In advance, however, we do not know who might be adversely affected by a nuclear waste management system; nor do we know his/her preferences. The last branch on the tree in Figure 6 represents the event that a particular person, m , with utility u_m suffers the l th health effect. The probability that this will occur, given that the accident and health effect scenarios have occurred, is depicted as r_{ijklm} .

Another important effect of a waste management system is its expense. Also on the last branch we have depicted a person m' as paying e_{lm}' (directly or indirectly) to support the system.

At the right most edge of a sequence of branches in the tree, we list l our ultimate concerns--the interpersonally calibrated utility that people would incur. In our example m incurs a utility loss u_m^l , and each of the m' incurs a loss of u_m' (e_{lm}').

Cast in this form, the analytical problem has horrendous informational requirements. We need to know the interpersonally calibrated utility functions of all those who might be significantly affected by the waste management system. Fortunately, a number of assumptions can be made that are reasonable in the waste management context and that greatly simplify the analysis.

(A1)* The health effects of system 1 are statistical in nature only.

By this we mean that the probability faced by any particular person of a serious health effect from the system is less than about 1/1000. (By way of reference, the probability that an American will die because of an automobile accident as opposed to other reasons is on the order of 1/100).

* Major assumptions in this paper will be indicated by A_.

(A2) The typical person's utility is a linear*, additive function of the various serious health effects, so long as the probabilities are statistical.

In ongoing work, R. Howard[†] has analyzed individual preferences for safety and safety expenditures in terms of more basic preferences concerning consumption, length of life, and risk attitude. Among other things, his analysis accounts for the fact that, typically, no amount of money will include individuals to take large health risks. On the other hand, Howard's analysis indicates that the amount of money a person will spend to avoid a statistical health risk is linear in the probability of the risk (e.g. if a person is to be consistent, doubling the probability doubles the amount that that person should be willing to spend). When this theory has been applied to the modelling of real preferences, individuals have generally agreed with these results of Howard's analysis.

Given A1 and A2 and the result of Howard's analysis, we can approximate m's interpersonally calibrated utility function by:

$$u_m = -w_m \left(\sum_{\ell}^L a_m^{\ell} p_{\ell} + a_m^{L+1} e_m \right) \quad , \quad (22)$$

where

w_m is m's interpersonally calibrated utility weight

p_{ℓ} is the probability of m's suffering the ℓ th health effect

The a 's represent m's tradeoffs among incurring the potential statistical health effects and making the dollar expenditure (e) to avoid them.

Note that A1 leads to the conclusion that individuals are risk neutral towards potential effects (i.e., their safety expenditures should be

* Changing any of the probabilities will change utility by a proportional amount.

[†] This work will be presented in the Proceedings of the 1978 Lawrence Symposium on Systems and Decision Sciences.

strictly proportional to the probabilities of health effects). This implies that public policy also should be risk neutral because the members of society are.

Our third simplifying assumption is that

- (A3) The health preferences of persons affected are equally likely to be any of those typical among current generations. (We assume here that the preferences of future generations among health outcomes are approximately the same as those of current generations. Improvements in medicine for preventing such outcomes should be considered in the probability assessment part of the analysis. The "time discounting" problem will be discussed below.)

Essentially, A3 allows us to use the average of the u_m 's in evaluating the social disutility of health effects, rather than having to estimate u_m for each of the potential victims. (In terms of the decision tree in Figure 6, we can simplify the end branches.)

From (22), the social disutility of h_ℓ occurrences of health effect ℓ becomes:

$$h_\ell/M \sum_{m=1}^M w_m a_m^\ell, \quad (23)$$

where M equals the U.S. population (or the number of people in a random sample of the U.S. population).

If we define

$$\bar{a}_\ell = \frac{1}{M} \sum_{m=1}^M \frac{w_m a_m^\ell}{M}, \quad (24)$$

then the social disutility becomes simply $h_\ell \bar{a}_\ell$. $\bar{a}_\ell$ is just a weighted average of citizens' tradeoff parameters a_m^ℓ , where the weights are the interpersonal comparison of utility coefficients, w_m .

Our fourth assumption is

- (A4) To a first approximation, the preferences of those who pay for the waste management system are the same as the preferences of the population at large, and the payers within a generation pay equal shares.

Under this assumption, the social disutility of an expenditure E by current generations* (relative to the disutility of health effects) is given approximately by

$$\sum_{m=1}^M w_m a_m^{L+1} (E/M) \triangleq E \bar{a}_{L+1}, \quad (25)$$

where $\bar{a}_{L+1}$ has the same weighted average interpretation as the $\bar{a}_L$ for health effects.

Our fifth assumption is

- (A5) In a society having a large middle class, such as the United States, the unweighted average of the a_m^L is probably a reasonable approximation to the weighted average in (24) and (25):

$$\frac{1}{M} \sum_{m=1}^M w_m a_m^L \approx \frac{1}{M} \sum_{m=1}^M a_m^L. \quad (26)$$

Under this assumption, we can approximate $\bar{a}_L$ by the unweighted average; and therefore, interpersonal comparison of utility no longer plays a direct role in the analytical calculations.

IV.B. Present versus Future

Comparing the utility of health effects and expenditures to current generations with the utility to future generations is a complex task. It is not simply a matter of discounting by the interest rate. Discounting applies to money, not utility. Furthermore, discounting by means of an

* E is total expenditures of current generations.

interest rate, only makes sense when all individuals have equilibrated their time preferences through the financial markets.

Utilitarians who make comparisons between the present and the future must consider (a) how well off future generations are likely to be compared with present generations, (b) how money not spent on safety today will benefit future generations, and (c) the general health risk environment in the future.

Note that considerations (a) and (c) tend to work in opposite directions. To the extent that future generations will live much better lives than we do, we might tend to discount their utility relative to ours. However, if they are leading much better lives, they are also likely to be living in a safer environment. Health risks that seem minute to us may be significant to them and so involve greater social disutility. For example, increasing the per capita chance of death by 1/1000 per year during the period of the Black Plague would be insignificant compared with making the same increase today.

A detailed analysis of the utility time discounting question is beyond the scope of this paper. For our purposes we will summarize the complex truth in terms of parameters, f_t , which will be used to express social tradeoffs between the statistical health risks to current generations and those to generations t years from now.

IV.C. Voluntary Versus Involuntary

How much should be spent to avoid occupational health risks versus the amount spent to avoid involuntary health risks to the public at large? From a utilitarian standpoint, the authors see no reason to distinguish the cases unless the workers facing the risks have markedly different preferences from those of the public. It is sometimes argued, however, that there is something worse about industrial health accidents involving the "innocent" public as opposed to industrial health accidents involving only employees, "who know the risks."

As with time discounting, it is beyond the scope of this study to examine this intricate question in detail. We simply let g be a parameter that discounts the (dis)utility of occupational health risks relative to involuntary health risks.

V INFERRING SOCIAL TRADEOFF PARAMETERS FROM EXPERT AND LAYMEN JUDGMENTS

The social utility function developed in the previous section has parameters $\bar{a}_h$ (representing the tradeoffs among various health effects and safety expenditures), g (representing the relative importance of occupational versus nonoccupational health effects), and f_t (comparing future versus present health effects). Because the utility function is linear and its overall scale is arbitrary, we need consider only the ratios of the $\bar{a}_h$ (e.g., dollars of safety expenditures per statistical quick death).

We now show how to estimate the parameters of Bayesian inference (Raiffa and Schlaiffer, 1967). Our "data" consists of judgments by experts and laymen about the parameters. The essence of Bayesian inference in this case is to represent both initial knowledge and accuracy of the judgments by means of probability distributions. These are called the "prior" and "likelihood" functions, respectively. The "data" that are observed (in this case, the judgments of experts and laymen) contain information about the parameters we are trying to estimate. The quality of the information is determined by the likelihood function. Given the data, we can calculate a probability distribution that represents our new state of knowledge. This calculation is determined by Bayes Rule, an elementary theorem in mathematics. The probability distribution we end up with is called the "posterior" distribution because it represents knowledge acquired after the data have been observed. Bayesian inference is just a quantitative theory of learning from experience.

The computations below are not intended to be recommendations concerning the values of the parameters. Our intention is only to demonstrate the Bayesian methodology.

V.A. Data

The Lawrence Livermore Laboratory (LLL) has identified five categories of health effects concerning nuclear waste management:

<u>Health Effects</u>	<u>ℓ</u>
Quick death	1
Fatal cancer	2
Nonfatal cancer	3
Mental subnormality	4
Sterility	5

For a more precise description of these events, see Attachment 1.

We will take $\ell = 6$ (and $\bar{a}_6$) to correspond to dollars of safety expenditures.

A contractor for LLL individually interviewed 58 citizens, asking them:*

- (i) The relative importance of the above health effects upon "unknown persons."
- (ii) How much "the Government should expend (in today's dollars) to avoid an instance (statistical) of one of these effects."
- (iii) The relative importance of occupational versus nonoccupational fatal cancers.
- (iv) The relative importance of fatal cancers occurring 30, 1000, and 10,000 years from now.

Thirteen of the citizens were national advisors. Their responses are given in Attachment 2.

The officials were not told to view the problem from the utilitarian perspective; therefore, we are not sure what considerations motivated their answers. However, we assume that utilitarian logic played a major role in their thinking. Of course, answering such general questions, even with counseling, is very difficult. For our purposes we regard the answers

* See Attachment 2 for reference.

as very "noisy" estimates of the "true" social tradeoff parameters that we would calculate if we had much more detailed information about Americans and their preferences.

We can also make use of the judgments of experts concerning the tradeoff between a statistical quick death and dollars spent to avoid such a death ($\bar{a}_1/\bar{a}_6$). We make no attempt here to carefully survey expert opinion, but use instead some willingness-to-pay (WTP) estimates that have been reported recently.*

Specifically, we use judgments by S. E. Rhoads (1978) that

a WTP proponent could argue that our current knowledge is sufficient to support a WTP figure two or three times DFE (discounted future earnings).

(DFE) can range from under \$100,000 to over \$400,000.

If we take DFE to range from \$75,000 to \$450,000, then two to three times DFE ranges from \$150,000 to \$1,350,000. For the purpose of constructing our prior function, we will suppose that 13 experts have reported 13 estimates ranging from \$150,000 to \$1,350,000, by increments of \$100,000.

We will also suppose that another 13 experts have reported the 13 numbers that appear in a sensitivity analysis carried out by R. Howard of the preferences of a sample person, 25 years of age. (See Attachment 3.)

V.B. Averaging

Why do we not just average the judgments and use these as our final estimates? The issue is one of whether averaging is justified or not in our inference problem. The usual justification for averaging is that the

* Note that an estimate of the mean willingness-to-pay is different from the utilitarian average, because the utilitarian average is weighted by the interpersonal comparison of utility factors, w_m . However, as we noted in assumption A5, in the United States which has a fairly straight distribution of income about the mean, the willingness-to-pay average is likely to be a reasonable approximation to the utilitarian average.

likelihood function describing data error stems from a normal or "bell-shaped" probability distribution, and that either the prior information is also normally distributed or it is nonexistent. These assumptions, however, are not particularly good when the parameters we are estimating have positive value, and when the mean of the data minus two standard deviations gives a negative value. In this case, the assumptions imply that there is a significant chance that the parameter has a negative value when we know for certain that it is positive.

Because of our uncertainty about the social tradeoff parameters, standard deviations of the judgments are likely to be large, in many cases, relative to the mean. Hence, the above assumptions about normality are questionable in our case. We therefore carry out the Bayesian analysis using more appropriate assumptions. We will still find that something similar to averaging is justified.

V.C. Prior Information

V.C. 1. Statistical Quick Deaths Versus Dollars

How should we model prior knowledge about the tradeoff between dollars and statistical quick deaths ($\bar{a}_1/\bar{a}_6$)? The authors propose constructing a prior distribution with the same mean and variance as the expert judgments.

What functional form should the distribution have? It seems reasonable to assume that the probability density should be distributed roughly like that in Figure 7. Only positive values of the tradeoff ratio have nonzero probability density, and the function is skewed to the right, reflecting the open endedness of "erring by too much."

A family of probability distributions having the form given in Figure 7 are the inverted-gamma-1's (Raiffa and Schlaifer, 1961, pp. 227-229):

$$p_{i\gamma 1}(x|n_1s) = \frac{e^{-s/x} \cdot (s/x)^{n+1}}{(n-1)!} \cdot \frac{1}{s} \quad \begin{matrix} x \geq 0 \\ n, s > 0 \end{matrix} \quad (27)$$

FIGURE 7 THE FORM OF THE PRIOR DISTRIBUTION ON THE TRADEOFF BETWEEN DOLLARS AND STATISTICAL DEATHS, P.

The mean of such a distribution is

$$\mu_1 = \frac{s}{n-1}$$

and the variance is

$$\mu_2 = \frac{s^2}{(n-1)^2 \cdot (n-2)} = \frac{\mu_1^2}{(n-2)}$$

The mode occurs at

$$\text{mod} = \frac{s}{n+1}$$

(This function is also known as Peurson's type V.)

In Figure 8, we have plotted an inverted-gamma-1 distribution with the same mean and variance as those of our hypothetical expert judgments. Note that almost all of the probability density lies between -1 and +3 standard deviations (s.d.) away from the mean. In other words, it is very improbable that the ratio lies outside this range according to our hypothetical expert opinion.

FIGURE 8 GRAPH OF $PIT^2 (\bar{x}_1/\bar{s}_1, n=3.07, s=3.76)$

V.C.2. Using the Expert Prior to Screen Response on the Dollar/Statistical Death Tradeoff

It is easy to become confused when evaluating how much the government should spend to avoid a statistical death. On the one hand, a person who thought that others besides the government should pay, might suggest an amount far below the average that individuals would be willing to spend in their own lives. On the other hand, a person who failed to consider the low probability, "statistical" assumption (A1) might suggest a very high amount that would reflect the enormous sums that individuals would typically demand before they would accept a relatively high probability of death. (For example, would the reader play Russian roulette for any amount of money?)

Two of the advisors in our sample estimated the value of avoiding a statistical death as being under \$25,000. Another advisor estimated the value at \$10 billion. In light of expert opinion, it seems very probable that these estimates either were not based on sufficient reflection or resulted from misconceptions about what was being asked.

What can be done about this problem of "spurious measurements"? The best solution is to re-interview the officials, if this is possible. If this is not possible, then their responses can be either dropped from the sample or "moved in" to the periphery of expert opinion.

The authors suggest using the -1 s.d. to +3 s.d. range about the mean of the expert prior as a way of screening out unsatisfactory responses. In estimating $\bar{a}_1/\bar{a}_0$, we will "move in" the unsatisfactory responses to the limits of this range [(\$.060 million, \$7.1 million) for our hypothetical expert prior].

V.C.3. Other Health Effects Versus Statistical Quick Deaths

The authors know of no expert opinion concerning the tradeoffs $\bar{a}_k/\bar{a}_1$, $k = 2, 3, 4, 5$. What distribution, p , should be used to represent prior knowledge of these parameters? Basically, we want a prior that reflects an unbiased, lack-of-information state. The estimates of the national advisors should carry maximum weight in determining the tradeoff estimate.

If we have no information about a positive ratio, x , then we are just as ignorant about $1/x$. Our prior, therefore, should have the same functional form in both cases. Because the two priors are mathematically related by change of variables, we get the following functional equation for p

$$p(x) = \frac{p(1/x)}{x^2} \quad (28)$$

Furthermore, our prior should have the same form if we merely change the units of the ratio by a factor, k . This gives us the functional equation

$$p(x) = k p(kx) \quad (29)$$

A prior having both of these "invariance" properties is

$$p(x) = \frac{1}{x}^* \quad (30)$$

We will use it in carrying out our inference of the intra-health tradeoffs and will see that it yields reasonable results (see Jaynes, 1968, pp. 227-240).

V.C.4. Occupational Versus Nonoccupational and Future Versus Present Health Effects

A priori, we know that a nonoccupational health effect to an unknown person is at least as bad as an occupational health effect. This means that the parameter g , which discounts occupational health effects, must lie in the range between 0 and 1. Lacking information about where it lies, we will use a "flat" prior distribution to represent our initial knowledge.

$$p(g) = \begin{cases} 1 & 0 \leq g \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (31)$$

*The probability density function $1/x$ is improper in the sense that it does not integrate to 1. However, it leads through Bayes Rule to posterior distributions that are proper.

Concerning future versus present health effects, the advisors were asked to compare fatal cancers occurring 30 years from now relative to cancers occurring 1000 and 10,000 years from now. Not surprisingly, the responses varied by several orders of magnitude.

If, however, we look at the average rates of discount per generation implied by the responses,* we find that they are much closer in value. If the 1000-year comparison is given by f_{1000} , then the implied average rate of discount per generation, f_{1000}^{25} , is given by:

$$f_{1000}^{25} = (f_{1000})^{1/40^\dagger} \quad (32)$$

The tables in Attachment 2 give the implied generational rates for both the 1000- and 10,000-year responses (f_{1000}^{25} , $f_{10,000}^{25}$). The authors propose to use the generational rates in carrying out the inference.

As for our prior knowledge, we know that the values of f_{1000}^{25} and $f_{10,000}^{25}$ are somewhere around one. We propose to use a flat prior about 1. The exact range of the prior will not affect the analysis, so long as the range includes all the estimates implied by the advisors' responses.

V.D. Likelihood of the Responses of the Advisors

Suppose we knew the value of one of the parameters, x , to be x_0 . Now consider a possible estimate of x_0 , $\tilde{x}$. What is the probability $p(\tilde{x}|x_0)$, that a thoughtful lay person will estimate x_0 to be $\tilde{x}$? Because we have no reason to suppose that the typical lay person's response will be biased above or below x_0 , we will assume that the mean of p is x_0 itself.

We know that the parameters f_{1000}^{25} and $f_{10,000}^{25}$ are on the order of 1. It seems reasonable to assume that the advisors' responses will also be close to 1 and have a relatively small standard deviation compared with

* We take a generation to be 25 years.

† 25 years x 40 generations = 1000 years.

the mean. Under these conditions, we can assume that p is approximately a normal distribution with mean, x_0 , but unknown standard deviation. We will estimate the standard deviation to be approximately the same as that of the advisors' responses.

More uncertainty surrounds the remaining parameters. The assumption that the standard deviation of the judgments will be small compared with the mean no longer seems reasonable. Instead, we shall assume that the likelihood functions are similar in form to the form of our expert prior, except that they have larger variances, reflecting greater uncertainty in the advisors' responses. Because of the difficulty in estimating the parameters, it is not unreasonable to suppose that the standard deviation of $\tilde{x}$ is roughly the same as x_0 itself.

Under these assumptions, we can approximate p by the exponential distribution:

$$p_e(\tilde{x} | 1/x_0^*) = 1/x_0 e^{-\tilde{x}/x_0}, \quad (33)$$

which is sketched in Figure 9.

FIGURE 9 GRAPH OF $p_e(\tilde{x} | 1/x_0^*)$

* The mean and standard deviations of $f_e(\tilde{x} | \lambda)$ are both equal to $1/\lambda$. In our problem, $\lambda = 1/x_0$.

V.E. Inference Results

V.E.1. Posterior Distributions (Except for f_{1000}^{25} and $f_{10,000}^{25}$)

Note that it is the inverse of the social tradeoff parameters that appears as the given quantity in the likelihood function. Hence, in order to apply Bayes Rule, we have to convert our prior distributions on the parameters to prior distributions on their inverses. This is an easy change of variable problem. The inverted-gamma-1 expert prior becomes simply a gamma-1 distribution (Raiffa and Schlaiffer, 1961, p. 227). The unbiased, lack of information priors on the intra-health effects keep the same functional form because of their invariance properties. The "flat" prior on g changes to a prior whose density is inversely proportional to the square of $1/g$ (letting $y = 1/g$, if $p(g)$ is flat on the interval $[0, 1]$, then $p(y) = y^{-2}$ on the interval $[1, \infty)$).

Applying Bayes Rule to the advisors' judgments, we get posterior distributions for the inverses of the social tradeoff parameters. For each of the three cases mentioned above, the posteriors turn out to be gamma-1 distributions (Raiffa and Schlaiffer, 1961, pp. 53-54). Changing variables to the uninverted parameters, we get the final posterior distributions which all have the inverted-gamma-1 form (the same as for our expert prior, but with different means and variances).

Table 2 gives the parameters of the final posterior distributions. S' is the sum of the advisors' responses; n' is the number of advisors (13 in our example); s is the sum of our hypothetical expert judgments about the dollars-statistical death tradeoff; n is the number of expert judgments (26).

Table 2
PARAMETERS (n'' and s'') OF THE POSTERIOR
INVERTED-GAMMA-1 DISTRIBUTIONS

$$p_{iyl}(x|n'', s'')$$

x	n''	s''
$\bar{a}_1/\bar{a}_6$	$n' + n$	$s' + s$
$\bar{a}_l/\bar{a}_1$ ($l = 2, \dots, 5$)	n'	s'
g	$n' - 1$	s'

Recall that the posterior mean (M_1), variance (M_2), and mode (mod) of the inverted-gamma-1 distribution are given by:

$$\begin{aligned}\mu_1 &= s''/n'' - 1 ; \\ \mu_2 &= \mu_1^2/(n'' - 2) ; \text{ and} \\ \text{mod} &= s''/(n'' + 1) .\end{aligned}\tag{34}$$

Note that if the number of responses, n' , is large, the posterior means and modes are both approximately equal to the means of the responses.

V.E.2. Posterior Distributions for f_{1000}^{25} and $f_{10,000}^{25}$

We have assumed a normal likelihood function and flat priors on both the f 's about 1. Under these assumptions Bayes Rule leads to a posterior distribution that is normal and has a mean equal to the mean of the advisors' responses. The standard deviation of the posterior is the same as that for the likelihood, which we have estimated to be approximately the standard deviation of the advisors' responses.

V.F. Numerical Results

Table 3 summarizes the numerical inference results using the judgments of Attachment 2.

Table 3

SUMMARY OF HYPOTHETICAL INFERENCE RESULTS

Social Tradeoff Parameter	Sum of Advisors' Responses	Posterior Distribution	Posterior Parameters		Posterior Mean	Posterior Standard Deviation
			n''	s''		
$\frac{\bar{a}_1}{\bar{a}_6}$ (million dollars per quick death)	\$28.205	Inverted- gamma-1	39	\$75.432	\$1.985	\$0.326
$\frac{\bar{a}_2}{\bar{a}_1}$ (quick deaths per fatal cancer)	40.72	↓	13	40.72	3.39	1.02
$\frac{\bar{a}_3}{\bar{a}_1}$ (quick deaths per nonfatal cancer)	8.19		↓	8.19	0.68	0.21
$\frac{\bar{a}_4}{\bar{a}_1}$ (quick deaths per mental subnormal)	25.52			25.52	2.13	0.64
$\frac{\bar{a}_5}{\bar{a}_1}$ (quick deaths per sterility)	1.91			1.91	0.16	0.05
g (occupational per nonoccupational)	5.28		12	5.28	0.48	0.15
f_{1000}^{25} *	12.30	Normal	Mean	S.D.	0.95	0.06
			0.95	0.06		
$f_{10,000}^{25}$ *	12.87	Normal	0.99	0.01	0.99	0.01

* See Subsection V.C.4.

REFERENCES

- Arrow, K. J., 1963, Social Choice and Individual Values, J. Wiley and Sons (Second Edition).
- Aumann, R. J. and Kurz, M., 1977, "Power in a Multi-Commodity Market," Israel Journal of Mathematics, Vol. 27, No. 3-4, 1977.
- Brock, H. W., 1978, "A New Theory of Social Justice Based on the Mathematical Theory of Games," in Game Theory and Political Science, Ed. P. Ordeshook, New York: New York University Press.
- Brock, H. W., 1978A, "The Shapley Value as a Tool for the Conceptual Unification of Economics, Politics, and Ethics," invited paper presented at the 1978 American Political Science Association Meetings, submitted to the American Political Science Review.
- Brock, H. W., 1978B, "Interpersonal Comparisons of Utility in Game Theory and Ethics: A Unified Account," invited paper presented at the Fourth International Conference on Game Theory, submitted to the International Journal of Game Theory.
- Brock, H. W., 1979, "Game Theoretic Insights into Ethics Based on a New Theory of Justice," Theory and Decision (forthcoming).
- Diamond, P., 1967, "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility: A Comment," Journal of Political Economy, 75.
- Harsanyi, J., 1953, "Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking," Journal of Political Economy, 61.
- Harsanyi, J., 1955, "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," Journal of Political Economy, 63.
- Harsanyi, J. C., 1963, "A Simplified Bargaining Model for the n-Person Cooperative Game," International Economic Review 4, pp. 194-220.
- Harsanyi, J. C., 1977, Rational Behavior and Bargaining Equilibrium in Games and Social Situations, Cambridge, England: Cambridge University Press.
- Howard, R., 1978, "Life and Death Decision Analysis," unpublished paper to be presented in the Proceedings of the 1978 Lawrence Symposium on Systems and Decision Sciences.

- Jaynes, E. T., 1968, "Prior Probabilities" in IEEE Transactions on Systems, Science, and Cybernetics, SSC-4, No. 3.
- Kaneko, M. and Nakamura, K., 1979, "The Nash Social Welfare Function," forthcoming in Econometrica.
- Marschak, J., 1950, "Rational Behavior, Uncertain Prospects, and Measurable Utility," Econometrica, 18.
- Nash, J. F., Jr., 1950, "The Bargaining Problem," Econometrica, 18.
- Raiffa, H. and Schlaiffer, R., 1961, Applied Statistical Decision Theory, M.I.T. Press.
- Rhoads, S. E., 1978, The Public Interest, 51, pp. 77, 83.
- Sen, A. K., 1970, Collective Choice and Social Welfare, San Francisco: Holden Day, Inc.
- Von Neumann, J. and Morgenstern, O., 1947, The Theory of Games and Economic Behavior (Second edition), Princeton: Princeton University Press.

Attachment 1

**EXAMPLES OF POSSIBLE HEALTH EFFECTS
OF EXPOSURE TO RADIATION**

Attachment 1

EXAMPLES OF POSSIBLE HEALTH EFFECTS OF EXPOSURE TO RADIATION

Quick death. Waste radiation causes quick death (within 30 days) to a person at age 40.

Fatal cancer. The victim gets a cancer at age 40, is somewhat ill (minor pain and discomfort) for 3 years, then very ill for one year, then dies.

Nonfatal cancer. The victim (age 40) is moderately ill with some pain, discomfort, and worry for 5 years, though functioning normally most of the time, then recovers completely and lives a normal life thereafter.

Mentally subnormal child. The radiation causes a genetic mutation in a parent who, as a result, has a child who is below normal mentally. All his/her life the child needs some special care, has to go to a special school, can hold only a paying job that is simple manual work; but can do most other routine life activities adequately. The child does not marry so there is no effect on future generations.

Sterility. The radiation causes a young woman to become permanently sterile so she can have no children, although she would like to have had children.

Source: Lathrop, J., "Development of Radiological Performance Objectives, Interim Results: Tradeoffs in Attitudes Toward Radioactive Waste," A-3, Lawrence Livermore Laboratory, July 1978

Attachment 2

**JUDGMENTS OF THIRTEEN NATIONAL ADVISORS ON SOCIAL TRADEOFFS
CONCERNING HEALTH EFFECTS AND PREVENTATIVE EXPENDITURES**

Table 2-1*

IMPORTANCE OF HEALTH EFFECTS RELATIVE TO QUICK DEATHS

<u>Subject Number</u>	<u>Quick Death</u>	<u>Fatal Cancer</u>	<u>Nonfatal Cancer</u>	<u>Mental Subnormality</u>	<u>Sterility</u>
1	1	2.00	0.100	14.00	0.010
2	1	12.00	3.000	2.00	1.400
3	1	14.30	3.700	1.20	0.033
4	1	1.00	0.017	0.25	0.170
5	1	0.22	0.011	0.33	0.022
6	1	0.85	0.110	1.54	0.077
7	1	1.92	0.500	2.31	0.019
8	1	1.22	4.3×10^{-3}	1.43	4.3×10^{-5}
9	1	0.07	4.0×10^{-3}	0.21	0.030
10	1	1.13	0.340	1.56	0.063
11	1	1.60	0.021	0.50	0.065
12	1	0.84	0.280	0.19	0.023
13	1	<u>3.00</u>	<u>0.100</u>	6.1×10^{-3}	9.1×10^{-5}
Sum		40.72	8.19	25.53	1.91
Mean		3.13	0.63	1.96	0.15
S.D.		4.53	1.23	3.70	0.38

* Calculated from *ibid.* p. 8.

Table 2-2

DOLLAR VALUES FOR PREVENTING A QUICK DEATH
30 YEARS* FROM NOW AS A RESULT OF A NONOCCUPATIONAL RISK

<u>Subject Number</u>	<u>Millions of Dollars</u>
1	0.100
2	0.020
3	0.007
4	0.360
5	0.810
6	0.360
7	150.000
8	10,000.000
9	3.500
10	0.330
11	0.380
12	8.600
13	1.000

* In the Bayesian Inference Analysis carried out in Section V, we distinguished between deaths happening in the near future, 1000 years from now, and those happening 10,000 years from now. However, we were not very precise about the time of near-term deaths because we wanted to avoid excessive detail in illustrating the methodology. In practice, the exact time of near-term statistical deaths is important in the utilitarian view only if there is significant variation in their disutility over the near-term time horizon.

Source: Lathrop, J., "Development of Radiological Performance Objectives, Interim Results: Tradeoffs in Attitudes Toward Radioactive Waste," p. 9, Lawrence Livermore Laboratory, July 1978

Table 2-3

RELATIVE IMPORTANCE OF NONOCCUPATIONAL
VERSUS OCCUPATIONAL HEALTH EFFECTS

Subject Number	Nonoccupational	Occupational
1	1	1.00
2	1	0.83
3	1	0.20
4	1	0.50
5	1	0.20
6	1	0.71
7	1	0.10
8	1	0.25
9	1	0.71
10	1	0.20
11	1	0.08
12	1	0.20
13	1	<u>0.30</u>
Sum		5.28
Mean		0.41
S.D.		0.31

Source: Calculated from *ibid.*, p. 11

Table 2-4

RELATIVE IMPORTANCE OF FATAL CANCERS
OCCURRING AT DIFFERENT TIMES

Subject Number	30 Years from Now	1000 Years from Now	$\left(\frac{.25}{1000}\right)^*$	10,000 Years from Now	$\left(\frac{.25}{10,000}\right)^*$
1	1	1.000	(1.000)	1.000	(1.000)
2	1	0.033	(0.918)	3×10^{-5}	(0.974)
3	1	0.010	(0.891)	10^{-5}	(0.972)
4	1	0.333	(0.973)	0.010	(0.989)
5	1	0.100	(0.944)	0.063	(0.993)
6	1	10.000	(1.059)	12.000	(1.006)
7	1	0.100	(0.944)	0.100	(0.994)
8	1	0.001	(0.841)	2×10^{-4}	(0.979)
9	1	0.920	(0.998)	0.830	(1.000)
10	1	1.000	(1.000)	1.000	(1.000)
11	1	0.030	(0.916)	0.020	(0.990)
12	1	0.100	(0.944)	0.071	(0.993)
13	1	0.004	(0.871)	4×10^{-5}	(0.975)
Sum			(12.30)		(12.87)
Mean			(0.95)		(0.99)
S.D.			(0.06)		(0.01)

* Implied average generational (25-year) discount rates. See Subsection V.C.4.

Source: Calculated from *ibid.*, p. 10

Attachment 3

**SMALL-RISK VALUES OF LIFE RESULTING FROM A SENSITIVITY ANALYSIS
OF THE PREFERENCES OF A HYPOTHETICAL 25-YEAR-OLD MALE**

Attachment 3

SMALL-RISK VALUES OF LIFE RESULTING FROM A SENSITIVITY ANALYSIS OF THE PREFERENCES OF A HYPOTHETICAL 25-YEAR-OLD MALE

Howard (1978) has developed a value model that relates an individual's small-risk value of life to his preferences for consumption and longevity of life and to his age and the interest rate. To illustrate the model, he has calculated the small-risk value of life for a hypothetical, 25-year-old male under certain assumptions about his preferences. He then varied the assumptions about the prevalences and age of the individual,* as well as the interest rate, to get the following set of 13 small-risk values of life:

Thousands of Dollars

529	2418	3622
1215	2430	3645
1277	2541	6648
1421	2671	6903
2157		

*The exact assumptions that led to the above small-risk values of life are not important in this case. We use the estimates only as hypothetical expert judgments about the mean willingness-to-pay in the United States to avoid small-risk hazards to life.